Accurate high-resolution 3D surface reconstruction and localisation using a wide-angle flat port underwater stereo camera

Towards autonomous ship hull inspection

Robert Schattschneider

A thesis submitted in partial fulfilment of the
requirements for the Degree of
Doctor of Philosophy

University of Canterbury, Christchurch, New Zealand

2014

Supervisor: Dr. Richard Green

Abstract

The goal of this thesis is to scan a ship hull with high 3D accuracy and resolution using an underwater stereo camera so as to enable the future autonomous detection of invasive biofouling organisms with autonomous underwater vehicles (AUVs). However, turbidity in most harbours necessitates being within a metre of the hull and thus requires ultra wide-angle camera lenses. But such ultra wide-angle lenses embedded in an underwater housing with a flat port lead to significant distance dependent image distortions. Prior research in this area has only considered narrower fields of view and so has not solved for the significant image distortions arising from wide-angle high resolution flat port underwater cameras. This thesis proposes a solution to modelling and calibrating the underwater camera for accurate 2D imaging and 3D reconstruction, and additionally demonstrates an accurate underwater real-time pose estimation system required for future ship hull relative AUV navigation.

In this thesis an ultra wide-angle, short-baseline stereo camera is used, which is embedded in a flat port underwater housing. Flat port underwater housings represent a cost efficient way to use arbitrary in-air cameras underwater. However, the flat port of the underwater housing is subject to light refraction and causes distance dependent distortion, which is particularly visible at the large angles of the ultra wide-angle stereo camera used. To incorporate the effects of refraction, the thesis uses the well-known and accurate physics-based refractive underwater camera model. In contrast to the perspective camera-based underwater camera model, the refractive underwater camera model accurately describes the distance dependency of distortion.

In the beginning of this thesis, the effects of refraction caused by a thick flat port underwater housing are summarised and extended. In this context, the fundamental magnification function is proposed, which enables the description of numerous known and also newly discovered effects. An additional quantitative analysis is carried out in which the importance to model the thickness of the port and the wavelength of light is revealed.

In refractive geometry with a thick flat port, refractive forward projection represents a fundamental operation and describes where a 3D object point is observed in a 2D camera image. Refractive forward projection is required in numerous applications, such as refractive calibration, bundle adjustment, simultaneous localisation and mapping (SLAM) or image restoration. Unlike perspective projection in air, this operation is non-linear and computationally more expensive. This thesis compares existing and proposes new refractive forward projection methods

and shows in contrast to previous research that refractive forward projection is efficient enough for real-time applications.

The thesis also investigates the impact of the port and the impact of the indices of refraction on the camera's projection and reconstruction accuracy. A novel investigation shows that the water pressure, water salinity, water temperature, air pressure and the wavelength of light significantly affect the projection and reconstruction accuracy of wide-angle flat port underwater stereo cameras and should not be neglected by standard refractive indices.

Moreover, this thesis proposes an accurate and efficient calibration method for thick flat port underwater stereo cameras. The proposed calibration method mainly achieves its high accuracy by the use of a significantly higher number of calibration images. In contrast to prior research, the computation of the reprojection error does not represent a bottleneck if the proposed refractive forward projection method is used. In this way, the calibration is similar to standard in-air camera calibration techniques and minimises the reprojection error.

In combination with the proposed more accurate indices of refraction and refractive calibration, the underwater reconstruction accuracy of the novel configuration of a wide-angle flat port underwater short baseline stereo camera is evaluated under real-world conditions. In this context, a method is proposed, which enables the evaluation of the accuracy of the reconstructed 3D object space.

Both chromatic aberration and pincushion distortion are effects of refraction and are particularly visible at the large angles of wide-angle underwater cameras. In order to obtained distortion-free images with minimised chromatic aberration to texturize reconstructed 3D ship hull surfaces, this thesis proposes accurate real-time methods to minimise chromatic aberration and to correct the distortion in the underwater camera images.

The refractive camera model is based on image coordinates of images, which are distortion-free in air. But these in-air undistorted images are strongly distorted in-water by refraction, particularly at the large angles of wide-angle flat port underwater cameras. Image correspondence in these images is difficult. For that reason, this thesis proposes pseudo rectified images in which these distortions are minimised. Moreover, an accurate and efficient representation of epipolar curves is presented, which enables, for example, real-time constrained correspondence search or dense stereo.

This thesis concludes with the demonstration of a pose estimation system for future ship hull relative navigation. The proposed pose estimation system is the first underwater SLAM and visual odometry system, which is based on the more accurate refractive underwater camera model. This thesis shows that the proposed pose estimation system is very accurate in a water tank experiment and efficiently works in real-time, and thus is superior to prior underwater SLAM research, which is based on the less accurate perspective camera-based underwater camera model.

Acknowledgements

I would like to thank the University of Canterbury, the Department of Mechanical Engineering and the Department of Computer Science and Software Engineering, which made my research for this PhD possible. I thank the University of Canterbury for the UC Doctoral Scholarship.

I also would like to thank the university for the technical equipment and the helpful support by the technicians, which all helped to make this research possible.

Moreover, I would also like to thank New Zealand and its people for the personal experiences in a new country and culture.

I especially thank my supervisor Dr Richard Green for the outstanding support during the last years. His excellent expertise, personal commitment, and valuable discussions strongly contributed to the quality of this thesis.

Most of all, I would like to thank my parents.

Contents

Acronyms

ACA	axial chromatic aberration
AHRS	attitude and heading reference system
AOV	angle of view
AUV	autonomous underwater vehicle
CA	chromatic aberration
CUDA	compute unified device architecture
DFOV	diagonal FOV
DIDSON	dual-frequency identification sonar
DOF	degree of freedom
DVL	Doppler velocity log
EKF	extended Kalman filter
FOG	fibre optic gyroscope
FOV	field of view
GPS	Global Positioning System
GPU	graphics processing unit
HAUV	hovering AUV
HFOV	horizontal FOV
IMU	inertial measurement unit
INS	inertial navigation system
KLT	Kanade–Lucas–Tomasi

LBL	long baseline
LCA	lateral chromatic aberration
LED	light-emitting diode
Lidar	light detection and ranging system
MEMS	microelectromechanical systems
MIT	Massachusetts Institute of Technology
MP	megapixel
MSIS	mechanically scanning imaging sonar
PeRL	Perceptual Robotics Laboratory
PMMA	polymethyl methacrylate
POI	plane of incidence
RANSAC	random sample consensus
RLG	ring laser gyroscope
RMS	root mean square
ROV	remotely operated vehicle
SBL	short baseline
SIFT	scale-invariant feature transform
SIMD	single instruction, multiple data
SLAM	simultaneous localisation and mapping
SSE	Streaming SIMD Extensions
SURF	speeded up robust features
SVP	single view point
ToF	time-of-flight
USBL	ultra-short baseline
VAN	visually augmented navigation

Notations

s	A scalar
$\boldsymbol{v}$	A vector
$\boldsymbol{M}$	A matrix
${}^{a}\boldsymbol{\pi}_{b}$	A pose $\boldsymbol{\pi}_b$ with respect to the coordinate system a
$\oplus$	The pose compounding operator (see section A.2)
$\ominus$	The inverse pose operator (see section A.2)
$f(x)$	A function and its argument
px	The pixel width or height, respectively, of a specific camera sensor

1. Introduction

The goal of this thesis is to scan a ship hull with high 3D accuracy and resolution using an underwater stereo camera so as to enable the future autonomous detection of invasive biofouling organisms with autonomous underwater vehicles (AUVs). However, turbidity in most harbours necessitates being within a metre of the hull and thus requires ultra wide-angle camera lenses. But such ultra wide-angle lenses embedded in an underwater housing with a flat port lead to significant distance dependent image distortions. Prior research in this area has only considered narrower fields of view and so has not solved for the significant image distortions arising from wide-angle high resolution flat port underwater cameras. This thesis proposes a solution to modelling and calibrating the underwater camera for accurate 2D imaging and 3D reconstruction, and additionally demonstrates an accurate underwater real-time pose estimation system required for future ship hull relative AUV navigation.

In this thesis an ultra wide-angle short-baseline stereo camera is used, which is embedded in a flat port underwater housing. Flat port underwater housings represent a cost efficient way to use arbitrary in-air cameras underwater. However, the flat port of the underwater housing is subject to light refraction and causes distance dependent distortion, which is particularly visible at the large angles of the ultra wide-angle stereo camera used. To incorporate the effects of refraction, the thesis uses the well-known and accurate physics-based refractive underwater camera model. In contrast to the perspective camera-based underwater camera model, the refractive underwater camera model accurately describes the distance dependency of distortion.

1.1. Autonomous ship hull inspection for the detection of invasive marine species

The threat of bio-diversity not only exists on land but also in waters. Crabs, fanworms, sea stars or clams as well as kelps and weeds belong to the most unwanted marine organisms in New Zealand. The Asian kelp, for example, is already present in New

Figure 1.1.: The Asian kelp is an invasive marine species, which has established in New Zealand's waters and displaces indigenous life forms [3].

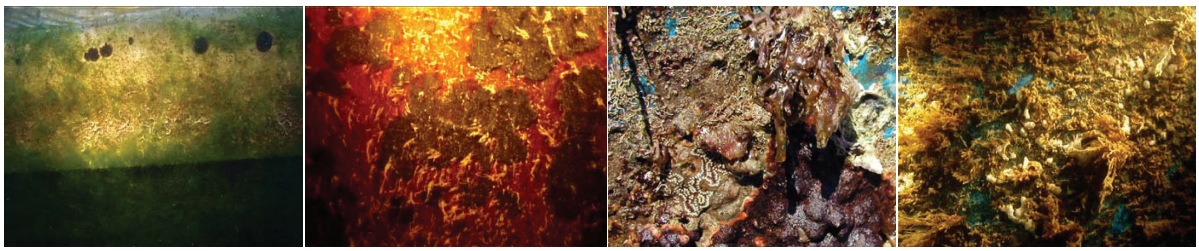

Figure 1.2.: Ship hulls are a major vector for the introduction of invasive marine species: The figure shows different degrees of biofouling from extensive to very heavy biofouling [4].

Zealand and is displacing the seabed's life forms (see Fig. 1.1). In addition to that, the reduced bio-diversity also endangers the fishing industry. Moreover, the congestion of water systems and the additional flow resistance of ships caused by fouling species influence the coastal infrastructure and reduce the ships' efficiency. Also the human health is endangered if introduced toxic species invade the food chain.

Shipping is the main reason for the introduction of non-indigenous aquatic life forms. The two main vectors are biofouling and ballast water. About three-quarter of the organisms are brought in by fouling on the hulls of vessels (see Fig. 1.2), whereas a further part is transported in the water intake systems, like ballast tanks or cooling water pipes. [5, 3, 6, 4, 7, 8, 9, 10]

Today, in order to avoid the invasion of foreign aquatic species, manual inspections are mostly carried out with pole cameras, divers or, since recently, with remotely operated vehicles (ROVs) [3, 11]. ROVs or AUVs significantly reduce the costs of in-water ship hull inspection [11]. Moreover, advanced future technologies for these underwater vehicles,

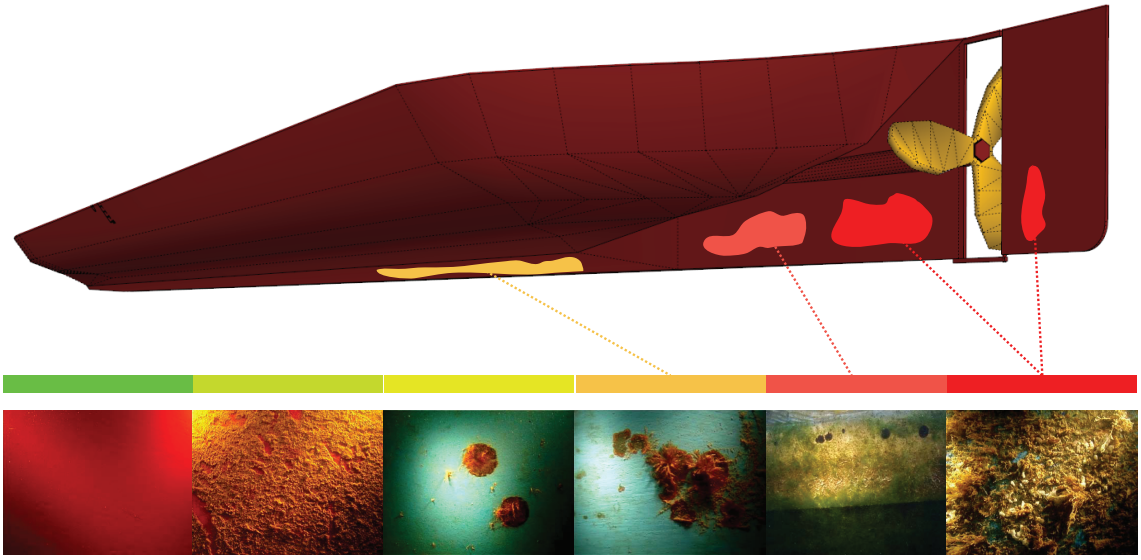

Figure 1.3.: The possible results of a proposed, future ship hull inspection system for the 3D reconstruction of the whole ship hull and the identification of the degree and level of fouling (images from [4])

such a ship hull localisation systems and image processing algorithms, will also enable the coverage of the whole hull, the generation of full 3D ship hull surfaces and the automatic detection of the degree and level of fouling (see Fig. 1.3). This thesis focuses on the accurate 3D reconstruction of the hull and a localisation system for such underwater vehicles.

1.2. Thesis overview

Chapter 2 This thesis begins with an introduction of state-of-the-art sensor technologies for underwater positioning and mapping, and then elaborates on more specialised technologies from industry and research for the inspection of ship hulls. Subsequently, the underwater camera technology, which is used in this thesis, is proposed and compare with prior research.

Chapter 3 In this thesis an underwater housing with a flat port is used. In addition to the flat port underwater housing, the thesis will also introduce alternative front port and lens-based underwater camera systems.

Chapter 4 The thesis proposes a fundamental magnification function for thick flat port underwater cameras, which enables the description of numerous known but also newly discovered effects of refraction.

Chapter 5 In contrast to the perspective projection in air, refractive forward projection is non-linear and computationally more expensive, but essential in real-time applications, which use, for example, fast refractive calibration, bundle adjustment, structure from motion, simultaneous localisation and mapping (SLAM) or image restoration. This thesis shows that in contrast to prior research refractive forward projection is actually possible in real-time applications.

Chapter 6 The thesis also investigates the impact of the port and the impact of the indices of refraction on the camera's projection and reconstruction accuracy. The proposed investigation shows that the water pressure, water salinity, water temperature, air pressure and the wavelength of light significantly affects the projection and reconstruction accuracy of wide-angle flat port underwater stereo cameras and should not be neglected by standard refractive index values. The investigation also reveals that in addition to the orientation and distance of the port, the thickness of the port should also be modelled for wide-angle underwater stereo cameras.

Chapter 7 This thesis then proposes an accurate and efficient calibration method for thick flat port underwater stereo cameras. The proposed calibration method leads to higher accuracies than prior calibration methods and is similar to standard in-air camera calibration techniques. In contrast to prior research, it is shown that the computation time of the reprojection error does not represent a bottleneck if the proposed efficient and accurate refractive forward projections method is used.

Chapter 8 The thesis also evaluates the underwater reconstruction accuracy of the novel configuration of a wide-angle flat port short baseline underwater stereo camera under real conditions. In this context, a method is also proposed which enables the evaluation of the 3D distortion of the reconstructed object space.

Chapter 9 The flat port of an underwater housing has a certain thickness. Hence, the incoming light is refracted at the port twice, that is, at the water-glass interface and the glass-air interface. In prior research, the refraction at the two interfaces was often simplified by single interface approximation, but the resulting error in 2D imaging and

3D reconstruction had not been evaluated. This thesis evaluates these errors, proposes a more accurate single interface approximation and shows that for high accuracies with wide-angle underwater cameras single interface approximation is not recommended.

Chapter 10 In order to texturize reconstructed ship hull models, the thesis then proposes accurate real-time methods to minimise chromatic aberration and to correct the distortion in the underwater camera images. Both chromatic aberration and pincushion distortion are effects of refraction and particularly visible at the large angles of the wide-angle underwater camera used.

Chapter 11 Methods for simplified image correspondence are proposed. Refractive camera models are based on in-air undistorted or rectified images, but these images are strongly distorted by refraction, particularly with wide-angle flat port underwater cameras. A method to minimise the distortion and to create pseudo rectified images is proposed. Moreover, the thesis proposes epipolar curves, which can be used in real-time, for example, for constrained correspondence search or dense stereo.

Chapter 12 The thesis concludes with the demonstration of a ship hull relative pose estimation system. The proposed pose estimation system is a novel underwater SLAM system, which incorporates the more accurate physics-based refractive underwater camera model. Moreover, the thesis also demonstrates that the novel solution of ship hull relative pose estimation based entirely on computer vision is possible. Furthermore, the proposed pose estimation system is very accurate and efficiently works in real-time, and thus is superior to prior underwater SLAM research, which is based on the less accurate perspective camera-based underwater camera model.

1.3. Research contributions

The many contributions of this thesis build upon each other to enable the two major goals of *accurate, high-resolution 3D surface reconstruction* and *ship hull relative localisation* with a high resolution, colour, wide-angle flat port underwater short baseline stereo camera. For such a wide-angle flat port underwater stereo camera, the following unprecedented evaluations and novel evaluation methods build upon each other to enable these successful proposed methods and functions:

1. Refractive magnification function Evaluation of the effects of refraction of a flat thick port underwater camera by a novel fundamental refractive magnification function, which enabled the descriptions of numerous known but also newly discovered effects. The evaluation also showed that lateral chromatic aberration and the thickness of the port should not be neglected in wide-angle underwater 2D imaging.

2. Fast and memory-efficient refractive forward projection methods Evaluation of the computationally efficiency of various refractive forward projection methods including a novel fast back projection-based refractive forward projection method and a novel memory efficient 2D lookup table-based refractive forward projection method. The evaluation revealed that refractive forward projection can be performed efficiently and does not represent a bottleneck in applications such as reprojection error-based calibration, bundle adjustment, SLAM or image restoration.

3. Identifying projection and reconstruction errors Evaluation of the projection and reconstruction errors caused by natural variations in the indices of refraction of water (salinity, temperature, pressure, wavelength), polymethyl methacrylate (PMMA) (temperature, wavelength) and air (temperature, pressure, altitude, humidity, CO_2, wavelength), and inaccurate thickness, orientation and distance of the port.

4. Accurate and fast calibration of the port pose and the dominating wavelengths Novel accurate and time-efficient calibration of the pose of the port of an underwater housing and of the dominating red and blue wavelengths. Here, the previously proposed fast back projection-based refractive forward projection method was used, which enabled the efficient minimisation of the reprojection error from multiple images.

5. Method to evaluate 3D distortion Evaluation of the 3D reconstruction accuracy of the proposed wide-angle flat port underwater stereo camera system, including a novel method to evaluated the distortion in the reconstructed underwater object space.

6. Improved single interface approximation Improvement of the single interface approximation including the evaluation of the imaging and reconstruction accuracy of the single interface approximation.

7. Real-time minimisation of lateral chromatic aberration and correction of refractive image distortion Two novel image restoration methods to minimise in real-time

lateral chromatic aberration and to correct refractive image distortion

8. Facilitated and fast correspondence search Novel pseudo undistorted and rectified images and novel real-time epipolar curves to facilitate (stereo) correspondence search for a wide-angle flat port underwater (stereo) camera.

9. Accurate and fast SLAM and visual odometry An underwater SLAM and visual odometry system, which uses the proposed accurate refractive camera model and proposed pseudo rectified images and efficient epipolar curves.

2. Underwater positioning and mapping for ship hull inspection

In this chapter, we firstly introduce common underwater sensor technologies for underwater positioning and mapping, then elaborate on ship hull inspection related technologies, and finally present the proposed positioning and mapping system for ship hull inspection.

2.1. Underwater positioning and mapping sensor technologies

2.1.1. Compass

Compasses enable absolute the measurement of the heading with the aid of the Earth's magnetic field [12]. The accuracy of compasses is 1-3°. Strong magnetic fields or structures with ferromagnetic materials considerably distort the measurements. Compensation methods are known to counteract static magnetic disturbances. However, no means are known to protect compasses against dynamic disturbances like electromagnetic fields caused by electrical motors. In [13], for example, a compass was used to determine the heading of an underwater vehicle within a harbour. For getting a distortion free signal, the motors of the vehicle had to be switched off and a distance to the metallic walls of at least 1-2 m had to be kept. Due to the interference with ferromagnetic materials, compasses are not useful for close-range ship hull inspection.

2.1.2. Depth sensor

The depth in-water, below the water surface, can be directly measured with pressure sensors. Strain gauge pressure sensors and quartz crystal pressure sensors represent the two most common technologies [12]. The depth is accurately derived from the water pressure [14]. The accuracy of strain gauge-based pressure sensors is $\sim$0.1 % and for

quartz crystal-based pressure sensors one order of magnitude higher (∼0.01 %).

2.1.3. IMU

Today's inertial measurement units (IMUs) are strapdown systems, which consist of rigidly mounted gyroscopes and accelerometers [15]. Depending on their accuracy, IMUs are deployed as inertial navigation systems (INSs) to determine the position and orientation, or as attitude and heading reference systems (AHRSs) to measure only the orientation. To determine the orientation and the position of an IMU, the sensor information of the gyroscopes and the accelerometers is often fused. Here, the gyroscopes are mainly used to calculate the orientation by integrating the measured angular velocity. The position of the IMU is computed by double integration of the acceleration measured. To eliminate the acceleration vector of the Earth's gravity, the orientation of the IMU needs to be known with high accuracy. Hence, the positioning accuracy of an IMU mainly depends on the accuracy of the gyroscope.

Fibre optic gyroscopes (FOGs) are the second most accurate gyroscopes. Using a large coil of optical fibre, the angular velocity of FOGs is measured by the interference of two light beams travelling in opposite directions. Ring laser gyroscopes (RLGs) are the most accurate gyroscopes and work similarly to FOGs, but use mirrors. RLGs and FOGs are part of INSs, but are expensive and are mostly integrated in military grade devices. Microelectromechanical systems (MEMS) gyroscopes are significantly cheaper and determine the angular velocity by measuring the Coriolis effect with vibrating parts. In general, MEMS IMUs are only eligible as AHRSs, since the accuracy of the gyroscopes is too low. The drift in position of standard MEMS IMUs after 60 s is more than 150 m and in combination with magnetometers more than 5 m [15].

In terms of our AUV ship hull inspection system, IMUs are interesting, as they can be easily mounted on the AUV and deliver additional sensor input. However, affordable MEMS IMUs are not useful as an INS and can be only used as an AHRS. Similar to compasses (section 2.1.1), MEMS IMUs with additional magnetometers are not eligible for the inspection of ferromagnetic ship hulls. All in all, AHRS only enable the absolute measurements of the direction of the gravity vector, since the estimated orientation around the gravity vector drifts over time. As the thesis will show, the drift in orientation of our proposed system is relatively low. Alternatively, determining the vertical direction is also possible with simple depth sensors.

2.1.4. Sonar-based systems

Sonar is a common technique used in underwater positioning and measurement systems. Sonar-based technologies measure the propagation time, the direction or the energy of sound. The resolution and attenuation of the received signal depends on the sound's frequency. The signal's resolution increases with higher frequencies while the maximum detection distance decreases.

2.1.4.1. Underwater acoustic positioning system

Long baseline (LBL) systems consist of a network of transponders and a hydrophone [12, 16, 17]. The transponder network is usually moored to the seafloor. The name of this acoustic positioning system originates from the large baseline, which ranges from a hundred to several thousand metres. The position of the hydrophone is determined by the round trip times of the acoustic signal between the hydrophone and each transponder using multilateration in a non-linear optimisation approach. Depending on the acoustic frequency used, LBLs can, on the one hand, achieve accuracies in the order of sub-centimetres (300 kHz) with high update rates of up to 10 Hz, or, on the other hand, long distances of up to 10 km with a lower precision between 0.1-10 m. The disadvantage of LBLs is the require time for the installation, calibration and maintenance, which is particularly necessary due to biofouling.

Short baseline (SBL) systems consist of hydrophones, which are often lowered from ships or floating platforms, and additionally consist of a transponder to which the position shall be determined [16, 17]. The distance between the hydrophone is 20-50 m. The positioning accuracy of SBLs is generally better than one metre.

Ultra-short baseline (USBL) systems are similar to SBLs and consist of a device, which incorporates an array of transducers. The transducer array is often attached to the bottom of a ship or a floating platform [12, 16, 17]. The distance between the transducers is 10 cm or less. The distance of the transponder to the array is determined by the signal runtime and its direction to the array by the phase shift measured in the array. USBLs have a much lower accuracy than LBLs due to the smaller baselines, but require a modest infrastructure compared to LBLs.

The disadvantage of all acoustic positioning systems is multipath interference, which occurs if the acoustic signal is reflected at the water surface or, for example, at the hull of a ship. Moreover, a direct line of sight between the transponders and transducers needs to be ensured. Additionally, acoustic positioning systems always consist of at least

two acoustic devices, so that an underwater vehicle to be positioned always depends on an external infrastructure. Another disadvantage lies in the small update rate when multiple participants have to share the same frequency bands.

2.1.4.2. Doppler Velocity Log

The Doppler velocity log (DVL) is an acoustic system, which measures the velocity with respect to the sea bottom, ship hulls or the water column [17] using the Doppler effect. The DVL consists of three to four beam transducers. The DVL is a dead-reckoning positioning system, where the position is calculated by integrating the measured velocity. The sample rate is up to 5 Hz and the accuracy in bottom velocity is about 0.4 % [18]. Limitations arise if the system operates too close to the tracked objects or in confined spaces [13].

2.1.4.3. Imaging sonars

A mechanically scanning imaging sonar (MSIS), as used in [13], scans the vertical 2D plane by a fan-shaped sonar beam and iteratively rotates the beam after each scan. In this way a 2D representation of the environment is generated. In general, mechanically scanned imaging sonars are relatively slow. In [13], the generation of a complete 360 ° scan took at least 6 s.

The dual-frequency identification sonar (DIDSON) is another imagine sonar, which provides near video frame rates of up to 21 Hz [19, 20, 21]. The DIDSON measures the echo of narrow, transmitted beams and is equipped with a focus system to adjust the depth of field. The acoustic intensities are measured in 96 directions and 512 distances with a horizontal field of view of 29°. The vertical aperture can be set to either 28° or 1° to use the DIDSON in an imaging or profiling mode for the creation of either 2D images or 3D point clouds with sub-decimetre accuracy [20]. Compared to camera-based imaging, the DIDSON can be deployed in turbid water and detects objects at distances between about 0.4 −90 m depending on the mode set. However, it suffers from a low resolution, accuracy, signal-to-noise ratio and spurious returns in cluttered and heavily biofouled environments, which make the automatic tracking of ship hulls more difficult and requires manual interventions [20, 22].

2.1.5. Vision-based systems

2.1.5.1. Monocular and stereo cameras

In contrast to acoustic underwater camera or imaging sonars, respectively (see section 2.1.4.3), optical cameras provide a rich set of information, such as a high resolution, a high accuracy, a high sampling rate and the ability to detect colours. On the other hand, the absorption and scattering of light rays by particles or water molecules [23] results in poor contrast, distance dependent and wavelength dependent attenuation of light and noise. The attenuation of sun light may require additional lighting, which can increase the backscattering of light by suspended particles in turbid water. Dynamic lighting conditions of shadow patterns cast by the water surface (caustic) represent another challenge for the feature tracking with monocular cameras, but can also be advantageous due to additional texture information if a stereo camera is used [24, 25, 26]. Water bubbles or aquatic snow represent some further challenges. In terms of ship hull inspection, the turbidity of water and the texture characteristic of the hull all impact the camera's ship hull inspection capability.

Monocular cameras and stereo cameras can be used to reconstruct the scene captured. Monocular cameras achieve 3D reconstruction by structure from motion, where a scene is captured from different views. In in-air structure from motion, monocular cameras are unable to determine the scale of the scene captured. To date, it has not been shown for monocular underwater cameras, yet, that a reconstruction of the scale is possible. On the other hand, stereo cameras enable the direct measurement of the scale and the reconstruction from a single camera pose.

2.1.5.2. Structured light, Time-of-flight and Lidar

Another approach for hull relative navigation was described in [27]. In this paper, three laser pointers and a camera were deployed. The camera measured the distances to the three laser points and thus also enabled the measurement of the orientation to the hull. This approach is, for example, useful for underwater vehicles to keep a fixed distance to ship hulls. A similar approach was proposed in [28].

The sensor system required for ship hull scanning needs to be able to track and to reconstruct ship hull surfaces. Active, structured light systems are advantageous to passive optical camera systems in reconstructing textureless areas and can increase the visibility in turbid water [29]. Various structured light systems using only few laser points [30, 27], laser lines [31] or more sophisticated binary projected patterns [29] are

known.

Only few publications reported on using the popular in-air structured light system, Kinect, underwater [32]. The experiments of Ozsvald [32], in which an in-air Kinect sensor was used to reconstruct submerged objects, showed the limited capability to reconstruct underwater objects at larger distances. In the following paragraphs, we evaluate numerous additional limitations of Kinect for use underwater.

Proposed evaluation of Kinect, Time-of-flight cameras and Lidar for ship hull inspection Kinect uses infrared light with a wavelength of $\lambda = 830\,\text{nm}$, which is strongly attenuated in water. Considering a submerged Kinect sensor embedded in an underwater housing and assuming for simplicity pure sea water and full light reflection at the ship hull, the maximum remaining light signal strength, S, using the minimal specified viewing distance of $z = 80\,\text{cm}$ results to $S = e^{-2.07 \cdot 2 \cdot z} = 3.6\%$ and drops to $\sim$0.03 % for $z = 2\,\text{m}$ ($\lambda = 800\,\text{nm}$) [23]. In all cases, S is very low and can probably not be sensed by Kinect. To solve this problem, a modification of the Kinect hardware would be necessary, in which the near infrared light source could be replaced by light of shorter wavelength, such as blue light.

The refraction of light at the port of the underwater housing leads to a distance dependent 3D distortion resulting in a distorted structured light pattern and a distortion of the reconstructed scene. It is unlikely that the in-hardware implemented pattern recognition and reconstruction algorithms of the Kinect sensor are able to handle these types of distortion or that a correction of the distortions is easily possible.

The horizontal FOV (HFOV) of 57° of the Kinect sensor is also relatively small. Light refraction at the flat port of an underwater housing reduces the (horizontal) field of view in water

$$\text{HFOV}_w = 2\sin^{-1}\left(\frac{n_a \sin(\frac{\text{HFOV}}{2})}{n_w}\right), \tag{2.1}$$

to $\sim$42° (Snell's law), where $n_a = 1$ and $n_w = 1.33$ denote the approximate indices of refraction of air and water.

Another clear disadvantage of the Kinect sensor is the minimum viewing distance, which is too large for ship hull tracking. The minimum viewing distance of the in-air Kinect sensor amounts to $\sim$80 cm. Apart from the strong light attenuation, as described above, the camera of the Kinect sensor needs to be close to the hull in order to recognise and track the textures of the hull. Under more challenging conditions, such

as turbid water, the maximum ship hull distance is even more restricted. The minimum viewing distance of the Kinect sensor is increased by the smaller FOV, and thus further aggravates the minimum distance problem.

Kinect is an active light system. In turbid water, it suffers from backscattered light emitted by the projector. To minimise backscattering, the projector needs to be placed far from the camera what causes shadows, and so parts of the camera image are not fully covered by the projected pattern. In classic underwater photography, two spot lights on the opposite sites of the camera are used to reduce this effect. Something similar, such as two pattern projectors, would be also necessary for the in-water Kinect case. Another problem for the pattern recognition algorithm would be the blur caused by the forward scattering of light.

Similar to Kinect, time-of-flight (ToF) cameras and light detection and ranging systems (Lidars) are generally based on infrared light and share similar disadvantages, such as strong near infrared light attenuation, backscattering and blur by forward scattering. The resolution of ToF system is also relatively low at about $200 \times 200\,\mathrm{px}$. Necessary modifications of these latter systems would be even more complex than modifications needed for the Kinect sensor.

2.1.6. SLAM

SLAM [33, 34] was firstly mentioned in the 1980s and has become popular over the last two decades. SLAM enables a system with adequate perception and processing capabilities to localise itself in an initially unknown environment. SLAM is often used in combination with dead-reckoning systems to bound the arising drift, which increases without SLAM with the distance travelled. To this end, fix landmarks of the environment are used as reference points. As a result, the drift of SLAM systems only increases with the distance to the starting point.

2.2. Ship hull inspection systems

2.2.1. Underwater vehicles

Underwater vehicles are employed for the inspection and maintenance of underwater structures, such as ship hulls. They are able to dive into much greater depths and stay longer underwater than is possible for divers. Underwater vehicles can be classified into ROVs and AUVs.

ROVs are connected with an umbilical or tether. The umbilical allows the energy supply, the control of the vehicle and the transmission of the sensor data to the water surface.

In comparison to ROVs, autonomous underwater vehicles carry their own power supply. Due to the missing tether, their operating range depends on their available energy. As the communication underwater is limited, these vehicles have to plan their motion autonomously.

ROV crawlers are equipped with wheels and are particularly appropriate to drive on large flat surfaces, such as ship hulls. The permanent contact to the hull surface is achieved by vertical thrusters pressing them against the hull. The LBV300-5 Crawler [35] and the Lamp Ray [30] are two commercial examples especially developed for ship hull inspection.

The Lamp Ray ROV crawler uses an acoustic system to analyse the condition of ship hulls [30]. In addition to the material assessment, its acoustic sensor system is also integrated into the ROVs' real-time tracking system. It allows the tracking of shell plates and lines on the hull as well as the mapping of the hull's shape. The localisation accuracy of the system is about 15 cm.

In [11], the feasibility of ROVs for the inspection of vessel biofouling was compared with the current inspection procedure carried out by divers. The evaluation ranked the ability of ROVs to inspect the hull and to assess niche areas as almost similar to the one of human divers. Also no significant difference in the inspection time could be identified. According to the evaluation, the large advantage of ROVs lies in the substantially lower costs and high flexibility if the vehicles need to be deployed at different operating sites.

2.2.2. Sonar-based ship hull inspection

Numerous publications on the sonar-based inspection of ship hulls exist [36, 37, 38, 39, 40, 27, 41, 42, 43, 44, 22, 45, 28, 11, 46].

2.2.2.1. AquaMap Shiphull

AquaMap Shiphull of Desert Star Systems [37, 46] is an acoustic positioning system for in-water ship hull inspection, which has been on the market for more than 10 years. It consists of four baseline stations, which are lowered from the deck of the ship, and one mobile station carried by a diver, ROV and AUV (see Fig. 2.1). This system requires a plan of the ship to register the positions of the baseline stations. AquaMap Shiphull can

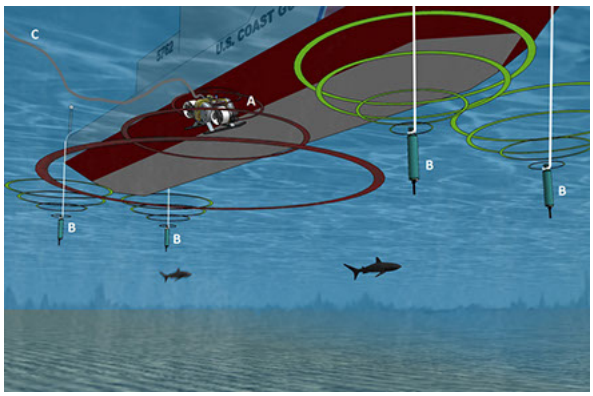

Figure 2.1.: AquaMap Shiphull of Desert Star Systems [46]: Positioning system for the in-water inspection of ship hulls for divers, ROVs or AUVs.

be, for example, used for visiting specific points on the hull or to track the mobile station to guarantee, for instance, the full coverage of the hull. The system can be also deployed for the inspection of other vertical structures, such as damns. The system requires an installation and calibration time, which is about 20 min according to the manufacturer. The position accuracy is typically 30 cm on the side and the bottom of the hull, and depends, amongst others, on how precisely the baseline stations can keep their positions, which is influenced, for example, by water currents. Furthermore, a clear line of sight between the mobile station and baseline stations needs to be ensured.

2.2.2.2. Hovering AUV

The hovering AUV (HAUV) project is a collaboration between, the Perceptual Robotics Laboratory (PeRL) at the University of Michigan, the Massachusetts Institute of Technology (MIT) and Bluefin Robotics, and is sponsored by the Office of Naval Research [20, 47]. This project is similar to our project and focuses on the inspection of ship hulls, however, with the main focus on the detection of mine-like objects and not biofouling. The HAUV is equipped with a comprehensive sensor suite that encompasses a DVL, a DIDSON, a monocular camera, an RLG and a depth sensor. The sensor data is fused in a graph-based SLAM framework [48], where the DIDSON and the monocular camera are used to extract features on the hull for drift compensated pose estimation. The camera has a limited (horizontal) field of view of about 45° in water and works with a low frequency of only 2 Hz [47]. Bolts, protrusions, holes, weld lines and biofouling on the hull represent some common features used. The distance of the HAUV to the hull is about 1 m.

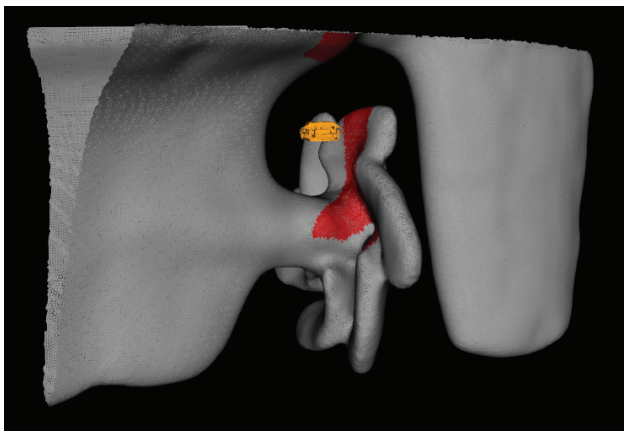

Figure 2.2.: HAUV - 3D reconstruction of the stern of a ship with an imaging sonar [20]. The resolution of the sonar is in the order of sub-decimetres.

Whereas the MIT group focuses on the acoustic mapping and view planning of the hull with the DIDSON [43, 44, 22], the PeRL group works on the visually augmented navigation (VAN) of the sensor suite with a monocular camera. The inspection of the hull for mine-like objects is mainly achieved by sonar, whereas the monocular camera primarily delivers additional constraints in the SLAM framework to limit the drift. The resolution of the DIDSON is in the order of sub-decimetres, which is enough for the detection of mine-like objects but insufficient for the detection of small biofouling organisms, which is the focus of this thesis. The reconstructed stern of a ship using the profiling mode of the DIDSON is shown in Fig. 2.2.

The PeRL group has only shown one monocular camera-based 3D reconstruction of the hull. The 3D photomosaic from 2009 from Kim and Eustice [49], as shown in Fig. 2.3, is the only vision-based reconstruction shown up to now. The reconstruction also only covers a part of the whole hull. The resolution of the 3D photomosaic is low and does not allow an assessment of the reconstruction quality achieved. In addition to that, as we will explain in more detail later in this thesis, Kim and Eustice use only the inaccurate perspective camera-based underwater camera model.

2.2.3. Stereo vision-based ship hull inspection

With respect to the sensor system and the field of application, the work of Negahdaripour and Firoozfam [24] is most similar to our research. Negahdaripour and Firoozfam use a stereo camera for the inspection of ship hulls. The pose of the stereo camera is estimated relative to only a flat projection of the hull. The pose estimation of Negahdaripour and

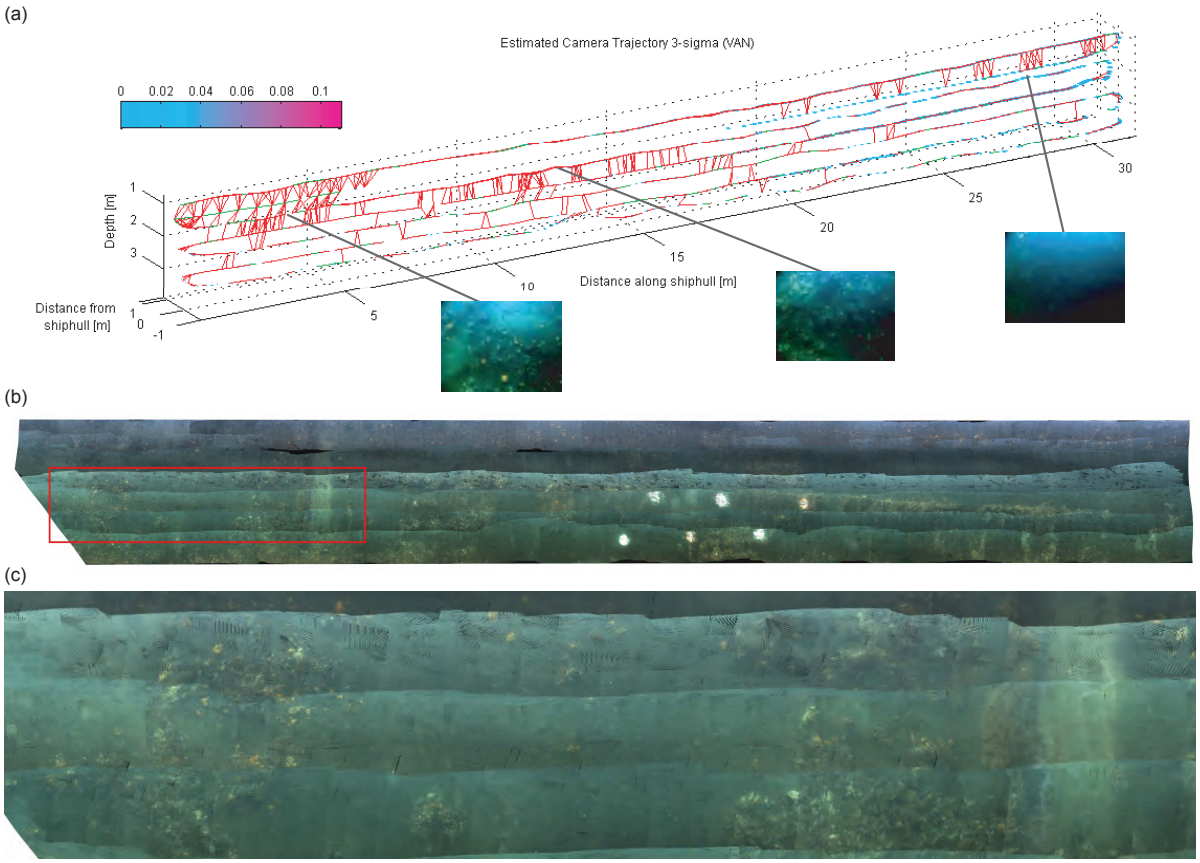

Figure 2.3.: VAN applied to the USS Saratoga aircraft carrier [49]: (a) Estimated 3D trajectory, (b) Textured map reconstruction, (c) Zoomed view

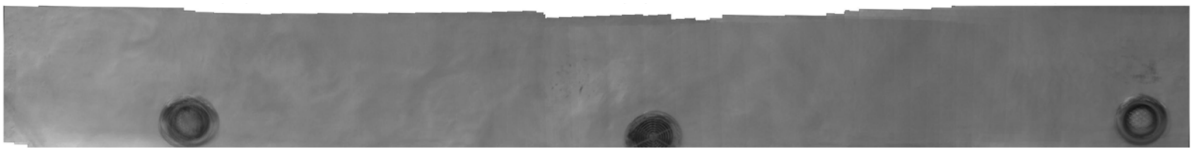

Figure 2.4.: Positioning of an ROV and generation of flat 2D photomosaic of the wall of a pool [24]

Firoozfam is a dead-reckoning system and does not compensate for arising drift. Like all the vision based systems presented in this chapter, the underwater camera model is based on the inaccurate perspective camera model. The pose estimation system was tested in a pool and in a sheltered dry-dock and demonstrated that a vision-based tracking is possible. An example image of the reconstructed flat 2D photomosaic of the pool is shown in Fig. 2.4. Negahdaripour and Firoozfam moved the camera about 7 m along the pool with an approximate wall distance of 40-120 cm. With this dead-reckoning system, the authors obtained an error of about 2 % with respect to the distance travelled.

2.3. Related vision-based projects

Finally, we would like to present two further interesting underwater vision-based projects, which are similar to ship hull inspection.

2.3.1. Damn inspection

The work of Ridao et al. [50] focused on the inspection of damns with a greyscale camera and multiple other sensors. The principle setup is depicted in Fig. 2.5. The inspection was carried out with an underwater vehicle and a surface buoy with a USBL to localise the vehicle. The vehicle itself was equipped with an MSIS, a DVL, a FOG, an Echo sounder and a monocular grayscale camera. The acoustic based sensors and the FOG ensured that the vehicle kept a frontal orientation and a constant distance to the wall to avoid distortions in the images captured. In an offline step, a large photomosaic with a resolution of 1 px/mm was generated, as also shown in Fig. 2.5. Again, the perspective camera-based underwater camera model was used, which is here sufficient, as the distance to the damn was not changed.

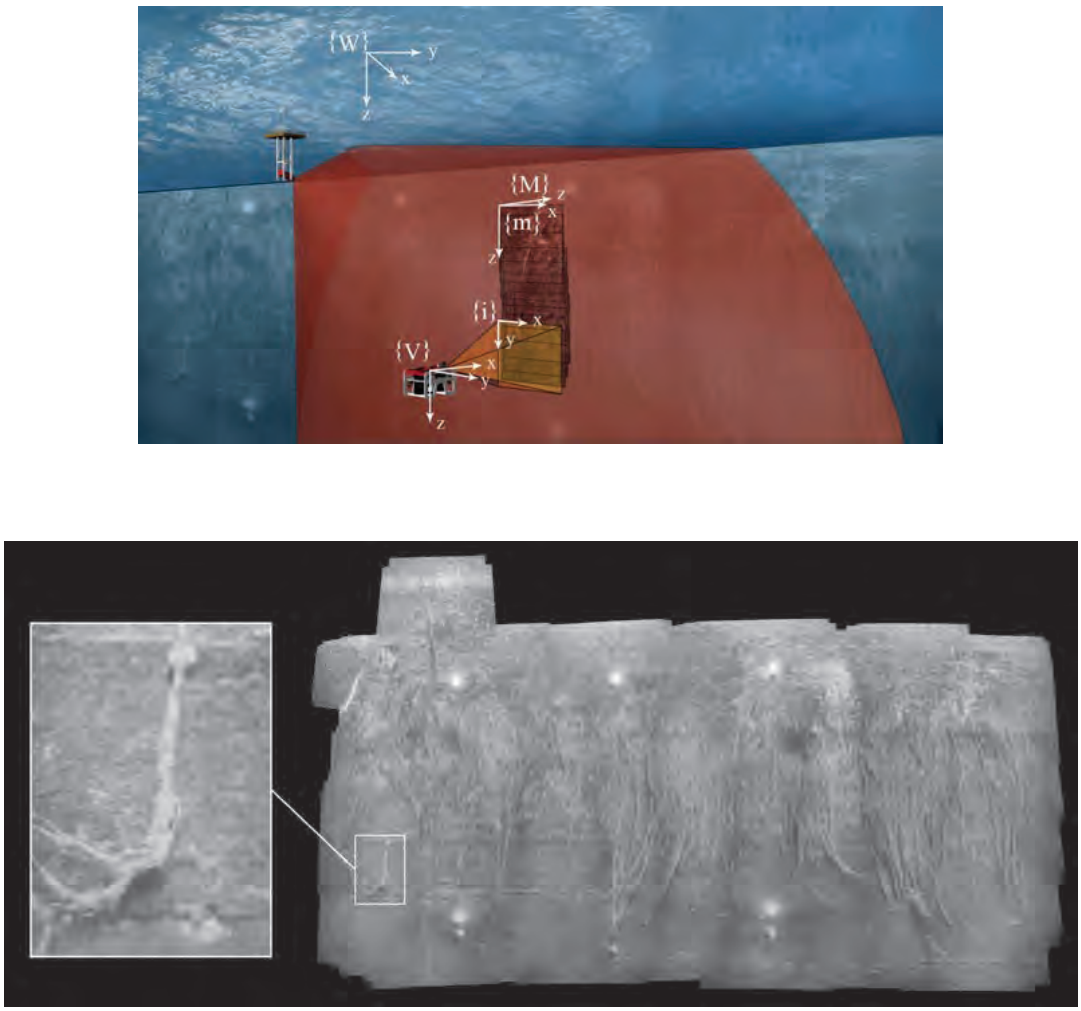

Figure 2.5.: Damn inspection [50]

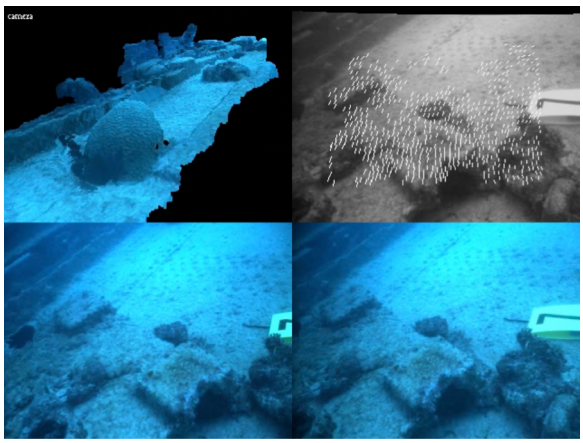

Figure 2.6.: Stereo vision-based reconstruction of a sunken barge [51]: Top left: Reconstructed surface from point clouds. Top right: 3D points for visual odometry. Bottom: stereo images

2.3.2. Reconstruction of a sunken barge

Hogue et al. [51] used an underwater stereo camera, which is, except for the smaller field of view, most similar to our underwater camera system used. The stereo camera was deployed for the mapping of a sunken barge, as shown in Fig. 2.6. The mapping system of Hogue et al. is based on visual odometry. Using dense stereo reconstruction, a point cloud was generated used to reconstruct the scene but also to determine the camera motion with respect to the last time step.

The drift of the map or of the motion was not limited by SLAM. Hogue et al. used the inaccurate perspective camera-based underwater camera model and observed a curvature of the recovered scene after several time steps. For that reason, the authors additionally employed an IMU to minimise this error.

2.4. Proposed system design

2.4.1. Wide-angle flat port underwater stereo camera

In this thesis, we will use a high resolution, colour, wide-angle flat port underwater short baseline stereo camera (see Fig. 2.7 and Table 2.1) in combination with SLAM. This system has several advantages, but also faces some challenges, which are solved in this thesis.

We have chosen optical cameras due to significantly more accurate information compared to imaging sonars. The resolution of visual cameras is significantly higher and so

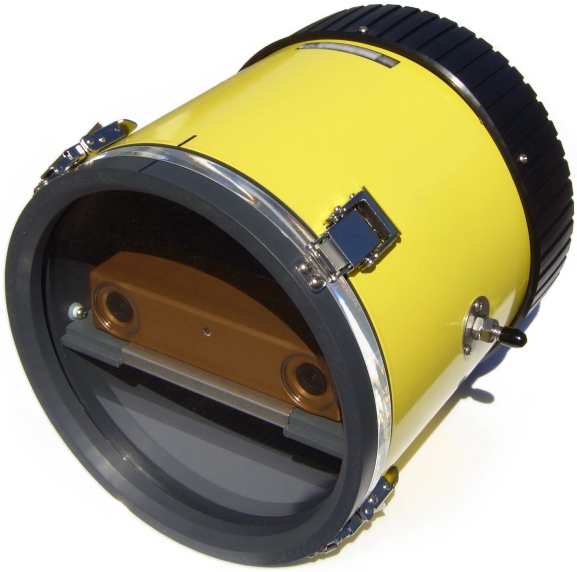

Figure 2.7.: The high resolution, colour, wide-angle (in water) flat port underwater short baseline stereo camera used in this research

Table 2.1.: Underwater camera specification

Stereo camera	Point Grey Bumblebee2 BB2-08S2C-25 [52]
Imager	Sony ICX204AK [53]
resolution (used)	1024 × 768 pixels
pixel number	0.79 MP
color	RGB, Bayer color filter array
format	4.8 mm × 3.6 mm
unit cell size	4.65 µm × 4.65 µm
Calibrated values	
baseline	12.07 cm
focal length	432.05 px, 2 mm
principle point	(512.85 px, 404.85 px)
HFOV (air)	99.68° (ultra-wide angle)
Lens material	unknown
Infrared cut filter	unknown
Port of underwater housing	
thickness	12 mm
diameter	19 cm
material	PMMA (acrylic)
HFOV (water)	70.15° (wide-angle)

more appropriate for the detection of small marine organisms. The colour information also helps to enable detecting biofouling. The higher frame rate of visual cameras also enables the real-time pose estimation required for the autonomous navigation of AUVs. In contrast to the VAN framework of the HAUV (section 2.2.2.2), the pose estimation can completely take place with only a single stereo camera, and does not require expensive navigation sensors, such as an RLG, a DVL or a DIDSON. In contrast to the narrow field of view of 45° in water of the monocular camera in the HAUV project, we use ultra wide-angle lenses with an ultra wide horizontal field of view of about 100° in air or a wide horizontal field of view of 70° in water, respectively. For a more detailed description of the terms *wide-angle* and *ultra wide-angle* see section 2.4.2. The wide-angle underwater camera allows a closer proximity to the hull, which is necessary in turbid water. Simultaneously, the vehicle can keep a relative high speed. For stable ship hull tracking, a narrow field of view would require a lower speed of the vehicle and a need to enforce smaller rotational velocities. A wide-angle underwater stereo camera is also beneficial for more accurate pose estimation and mapping, as a point on the hull is sensed from many more viewing angles. In comparison to the acoustic positioning system of Ridao et al. [50], our mapping and pose estimation system is completely independent from any external infrastructure. Additionally, as shown by Ridao et al. [50] and Hogue et al. [51], cameras enable the high resolution 3D mapping of underwater structures. In contrast to the 2D stereo vision-based ship hull reconstruction of Negahdaripour and Firoozfam [24], we focus in the thesis on the 3D reconstruction of the ship hull. Moreover, our mapping and localisation will be based on SLAM to bound the drift and to increase the accuracy in 3D reconstruction. In contrast to all prior SLAM-based underwater mapping research, this thesis will incorporate the more accurate physics-based underwater camera model, which significantly increases the reconstruction accuracy in 2D imaging, 3D reconstruction and 3D pose estimation.

2.4.2. Discussion of the terms wide-angle and ultra wide-angle

The term angle of view (AOV) describes the angular extent of the scene captured by a camera. In classic photography, the term AOV is common and interchangeable with the more general term field of view (FOV). Unless otherwise specified, the AOV refers to the diagonal FOV (DFOV). The AOV depends on the format of the image sensor and the focal length. In photography, it is stated in mm instead of °. The unit mm originates from the focal length of analogue 35 mm film cameras or digital full frame cameras with a film or sensor format of 36 mm × 24 mm.

There is no clear definition of wide-angle and ultra wide-angle lenses. Wide-angle lenses often have focal lengths of 24 mm-35 mm, whereas ultra-wide angle lenses with (fixed) focal lengths of 14 mm-20 mm are currently sold [54, 55]. These focal lengths correspond to a HFOV of 54°-74° for wide-angle cameras or 84°-104° for ultra wide-angle cameras, respectively. Here, the HFOV in ° was calculated by the following commonly known formula:

$$\text{HFOV} = \arctan\left(\frac{w}{2f}\right)\frac{360^\circ}{\pi}, \tag{2.2}$$

where w denotes the width of the image sensor and f the focal length.

Applied to our camera, this means that a HFOV of about 100° is ultra-wide in air. In-water however, the HFOV decreases to 70°, which corresponds to a wide-angle HFOV (see (2.1)). For that reason, we call our underwater camera a wide-angle underwater camera.

3. Front port systems and wet-lenses

3.1. Introduction

Different optical systems for underwater imaging exist. The solutions include front port systems in connection with underwater housings, which embed usual in-air cameras, or alternatively, waterproof cameras with wet-lenses, which are specially designed for the use in water [56]. In the following sections, we give an overview of different optical systems for in-water imaging. We introduce the flat port, the dome port, the Ivanoff corrector and a well known wet lens product series and discuss their characteristics.

3.2. Flat port

Flat ports of underwater housings are front port systems with advantages and disadvantages. Disadvantages consist in a narrowed field of view, chromatic aberration and pincushion distortion. As chromatic aberration and pincushion distortion are especially visible at larger angles, the use of flat ports is generally only recommended for lenses with a narrow field of view, such as macro or telephoto lenses [57].

Compared with alternative in-water optical systems, advantages of plane ports consist in low cost manufacturing and mounting, the suitability for short-baseline stereo vision, the additional magnification at macro and telephoto shots, the small curvature of field and the less complex optical geometry. Due to the simple shape, the casting of optical grade acrylic ports is significantly easier than the manufacturing of spherical dome ports or of specific underwater lenses. The alignment of flat ports with respect to the camera is not crucial, as it can be easily calibrated. In contrast to dome ports, to the Ivanoff Corrector or to wet lenses, multiple cameras can cost-efficiently share a common port. Compared with dome ports, the minimum baseline required in close-range stereo vision is not limited. For dome ports large radii are recommended to reduce the strong curvature of field and to increase the apparent object distance. But this limits the minimum baseline for stereo vision, though. Furthermore, the pincushion distortion of flat ports

can be beneficially deployed to partially compensate for the undesirable barrel distortion of ultra wide-angle lenses. As we will show in this chapter, the simple geometry of flat ports allows a straight forward calibration of the port's pose (chapter 7), the correction of unwanted effects in 2D imaging (chapter 10) and accurate 3D reconstruction (chapter 8).

3.3. Dome port

The dome port is a front port system, whose centre of curvature ideally coincides with the camera's centre of projection. Dome ports preserve the field of view underwater, reduce chromatic aberration and image distortion and can withstand high pressure [58]. If optimally aligned, dome port systems can be well approximated by the comparatively simple and well-known standard perspective camera model, and thus, standard techniques used for camera's in air [59] can be exploited. Using the standard lens distortion model [60], sufficient correction of image distortion is possible with imperfect camera-to-dome port alignment, but also under optimal calibration conditions, the remaining distortion still depends on object distance [59].

Further disadvantage of dome ports are the strong reduction of the apparent object distance, the strong curvature of field [58, 61], slightly darker images due to a smaller f-number [58] and cost intensive manufacturing and alignment [58].

An undesirable effect for close-range vision is that objects, which are actually located at infinity, are perceived by the camera at a close distance, which corresponds to only about three times the radius of the dome [61]. Hence, the closest object distance is limited by the shortest available focal length of the camera. Close-up lenses are used to reduce this limitation but are subject to degradation in image quality.

The curvature of field of a dome port is another limiting factor. The lens of a camera designed for the use in air, cannot sharply focus the curved virtual image on the camera's flat image plane. The loss in sharpness at larger radial distances is particularly visible at larger apertures with a reduced depth of field. This means a significant limitation for high resolution wide angle stereo.

In order to increase the apparent object distance or to reduce the field of curvature, large dome radii are recommended, in general. However, in short baseline close-range stereo vision, the size of the dome ports is limited by the distance between the cameras.

3.4. Ivanoff Corrector

The Ivanoff Corrector is a more sophisticated front port system consisting of a plano-convex lens being in contact with water and a bi-convex lens placed in front of the camera [62, 63]. The virtual image created by the Ivanoff Corrector coincides with the object so that the magnification and the field of view are preserved and images stay distortion free.

The Ivanoff Corrector is not free of disadvantages, either. For a better correction of chromatic aberration, the corrector needs to be extended by further lens elements resulting in increased costs. In comparison to flat ports, the costs for manufacturing and alignment are higher, too. Another cost factor is that the Ivanoff corrector is lens specific, and thus, cannot be used with arbitrary cameras lenses. The optical geometry also depends on the index of refraction of water and changes, for instance, if the corrector is used in fresh or salt water. The limited field of view represents another disadvantage. Since the camera has a certain distance to the plano-convex lens, the diameter of the plano-convex lens grows relatively quickly with greater field of views, and is therefore not useful for wide angle stereo. Wakimoto [58] also criticised that the corrector was only designed on the basis of Gaussian optics, which describe the optical system only for small angles.

3.5. Wet lenses

Waterproof cameras with wet lenses specially designed for the use in-water, represent the most accurate but also the most expensive solution in underwater imaging. Nikonos cameras with corresponding underwater (UW) Nikkor lenses were over a long time one of the best known underwater imaging solutions in underwater photography and photogrammetry [58, 56, 64]. Nikonos products were manufactured from the 1960s for about 40 years and are popular until today. Various underwater Nikkor lenses exist, which consists of a flat or dome shaped water interface. Depending on this interface, different alternative correcting lens systems are employed to achieve an image quality comparable to the one in air with a wide field-of-view, low distortion, low chromatic aberration, a high f-number and low curvature of field. Even though, the cameras or lenses were appreciated due to their superior characteristics, the series was a too specialised and expensive niche product and, thus, discontinued in 2001 [65]. In general, wet lenses for stereo vision exceed the cost of flat port solutions significantly.

4. Effects of refraction

4.1. Introduction

The use of a flat port in an underwater camera has numerous effects on 2D imaging. Some of these effects were mentioned in prior literature, but have not been summarised, yet. In this chapter, these refractive effects are summarised and background literature is clarified. Moreover, a fundamental magnification function is proposed, which enables the description of many effects in flat refractive geometry in a short form. This fundamental magnification function was also used to discover numerous new effects, which will be presented in this chapter. In addition to that, it will be shown that chromatic aberration and the thickness of the port of the underwater camera should not be neglected in wide-angle flat port underwater camera vision.

4.2. Related literature

The effects of refraction on image perception with one or two flat refractive interfaces have not been fully described yet. Recent literature is still addressing the problem of how certain parameters affect the image perceived behind a thin flat interface or a thicker window of a pool, aquarium or underwater housing.

In 1956, Ivanoff et al. [62, 63] mentioned in the context of their proposed correcting underwater lens how flat ports (windows) of camera housings change the optics of underwater cameras. Flat ports cause objects under water to appear about 4/3 times closer and larger and reduce the field of view and depth of view. The factor 4/3 is here equivalent to the ratio of the index of refraction of water n_w and the index of refraction of air w_a. The focal length of the camera needs to be reduced by about n_w/w_a to focus the apparent closer objects [63, 61]. In contrast, due to the reduced field of view and image magnification, images appear as if they were taken by a camera with a n_w/w_a larger focal lengths. Moreover, flat windows reduce the image quality by introducing chromatic aberration, astigmatism and distortion. The properties of flat ports as described

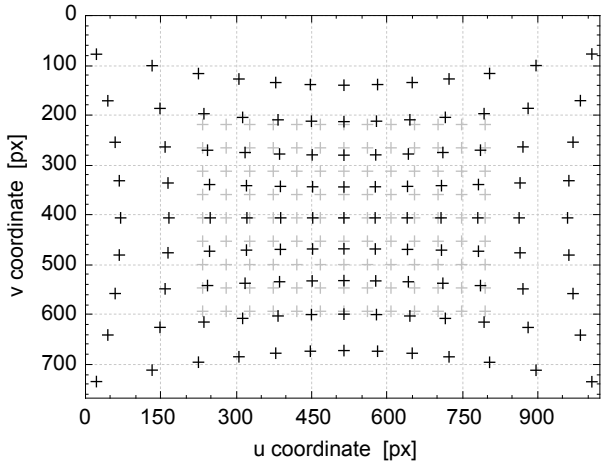

Figure 4.1.: Magnification effect and pincushion distortion by the refraction of light at a flat port (black crosses) in comparison to a perspective projection (grey crosses) of a point grid

by Ivanoff et al. and often enumerated in various literature are based on Gaussian optics and thus are only approximations for small angles and large object distances. Numerous effects at larger angles cannot be described. We propose a function that can exactly calculate refractive image magnification.

In his work on the design of underwater camera lenses from 1967, Wakimoto [58] described that flat ports produce a curved virtual image, so that flat surfaces cannot be exactly focused and appear blurry in the camera image at large radial distances. In comparison to the curvature of field of dome ports, the curvature of field of flat ports is small, however, as can be seen in [61]. Moreover, Wakimoto explains the occurrence of pincushion distortion, where the magnification increases with increasing angles. Pincushion distortion was illustrated in [66] and is depicted in Fig. 4.1. In addition to that, the effect of chromatic aberration was described, which is visible as red coloured fringes in the inside and blue coloured fringes on the outside of sharp edges. Wakimoto also mentions that at large object distances, the port thickness can be neglected. In this section, we demonstrate that the port thickness needs to be taken into account for close object distances, though.

In 2000, Kwon and Lindley [67] demonstrate that pincushion distortion cannot be modelled as a linear function of the image-plane coordinates alone, but also depends on the position of the object point in 3D space.

In 2008, Treibitz et al. [68] neglected the thickness of the port of their underwater housing, since only a little effect on image distortion was expected. It was demonstrated

that the radial pixel shift (radial distortion) increases with decreasing interface distance and, thus, especially occurs in underwater housings. The theoretical case of a single view point (SVP) model was studied, in which the interface is exactly located on the camera's centre of projection so that all virtual focal points coincide. For the SVP case, a novel accurate radial distortion correction function was presented, which is object distance independent. Treibitz et al. have not mentioned, however, that the SVP model exists only for small angles, as we will show. For larger angles, which are determined by the total internal refraction within the port, multiple virtual focal points occur. Therefore, the still correct physics-based radial distortion correction function should not be directly related to the SVP model. Moreover, we show that the port thickness cannot be neglected any longer in the case of high-resolution, wide-angle underwater cameras and close object distances.

In the same year, Kunz and Singh [59] demonstrated the inadequacy of the often used standard perspective camera model and the standard lens distortion model to correct the image distortion of a thick port. It was also observed that the rotation of the port has a stronger effect on distortion than its displacement and cannot be appropriately approximated by the decentering/tangential distortion terms of the standard lens distortion model. A visualisation of how an object is distorted by a tilted port was not given, though, and is completed by us in section 4.3.

In 2009, Chari and Sturm [69] established analogies between perspective projection and single interface refractive projection. Based on the derived refractive projection matrix, it was shown that a line in object space is projected to a quartic curve in image space. In terms of epipolar geometry, the refractive fundamental matrix was deduced, which relates the image points of an object point from two different camera views by the epipolar quartic curve. The main focus of Chari and Sturm is on theory and does contain actual visualisations of the epipolar quartic curve.

Gedge et al. [70] computed epipolar quartic curves (refractive epipolar curve) by piecewise linear approximation, visualised and compared them with perspective epipolar lines and investigated the pixel distance between the curves. Unfortunately, only small image sections of the refractive epipolar curves were shown, but not their location over the whole image, which makes an understanding of the characteristics of the epipolar curves difficult. To clarify characteristics of epipolar curves over the whole image, a supplement figure is presented by us in section 4.3.

In 2012, Treibitz et al. [71] showed that also increasing object distance results in increasing radial distortion, which is similar to the previously presented case of a very close

interface [68]. In both cases, the system converges against a SVP model for which the distortion can be modelled independently from the object distance. Moreover, Treibitz et al. extended their physics-based radial distortion correction function [68] for arbitrary angles and port distances by incorporating the object distance as an additional parameter. Using this correction function, an accurate method was demonstrated to reduce the distortion of objects at known distances. Our work is strongly related to this work of Treibitz et al.. Their radial distortion correction function is focused on a single interface system, whereas we extend it for thick port systems. Additionally, we deduce the fundamental magnification function, which is closely related to the radial distortion correction function of Treibitz et al. and highlight its importance for the understanding of numerous effects of flat refractive geometry.

The work of Yau et al. [72] is one of the few recent publications that account for the effects of wavelength dependent refraction (dispersion). The effect that different coloured light is perceived at different image coordinates was exploited for triangulation-based calibration.

4.3. Proposed visualisation and clarifications of background literature

Supplementary to existing literature, we would like to briefly present the visualisations of the refractive epipolar curve over the whole image [69, 70], the image distortion caused by a non-parallel port [59] and the limitation of the validity of the SVP model [68, 71]. All findings result from the physics-based refractive camera model.

In classic stereo vision, the epipolar geometry is used as a constraint for efficient stereo matching. The corresponding second image point in the second image is located on the straight epipolar line. It is interesting to see, how far the classic epipolar constraint is applicable in flat refractive geometry. The underwater stereo camera setup and the simulated camera images for the epipolar geometry are depicted in Fig. 4.2. To exemplary visualise the refractive epipolar curves, points from 9 different regions in the left camera image (Fig. 4.2b) were back-projected to their corresponding light rays in object space, and, subsequently, multiple points located on the light rays were forward-projected onto the right camera image. The resulting projections in the right image are visualised in Fig. 4.2c. The figure shows that the curvature of the epipolar curve is especially strong in the upper right and lower right region of the right camera image (curves 3 and 9). Here, a maximum change in the vertical coordinates of up to about 30 px is observable

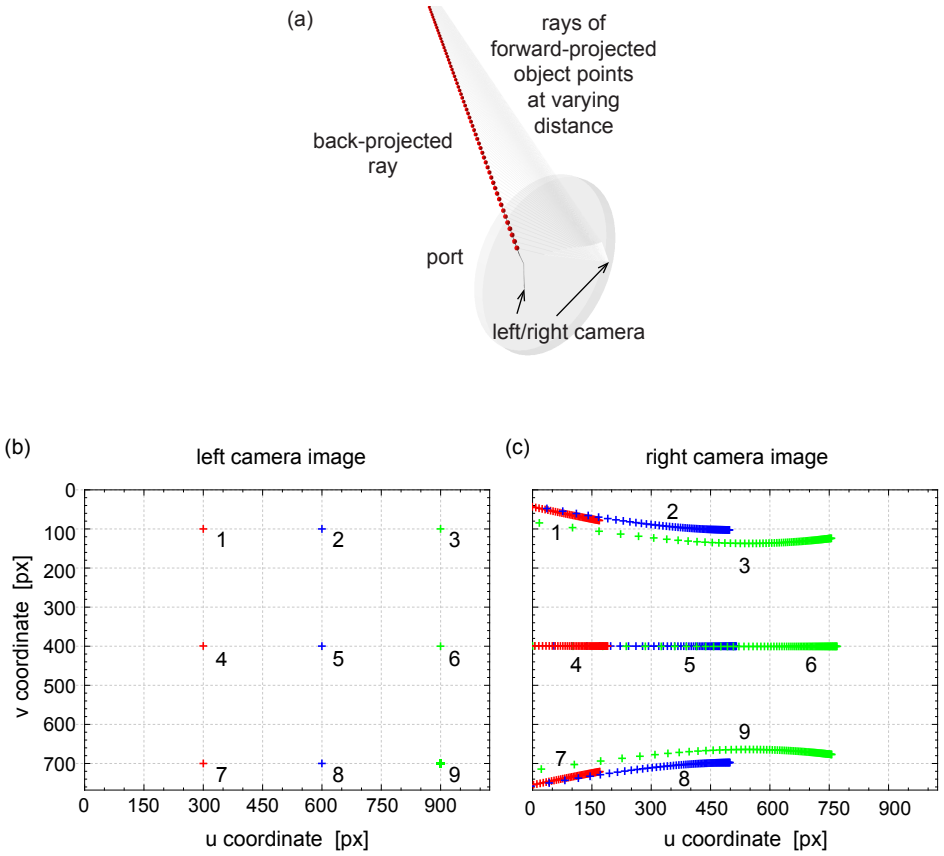

Figure 4.2.: Refractive epipolar curves of an underwater stereo camera with a flat, thick port: Figure (a) shows the 3D visualisation of the epipolar constraint, which describes where an object point perceived by the left camera (point 3) with unknown distance could be perceived in the right camera image. Figure (b) and (c) depict the left and right camera images, which show for 9 different image points in the left camera image the resulting refractive epipolar curves in the right camera image.

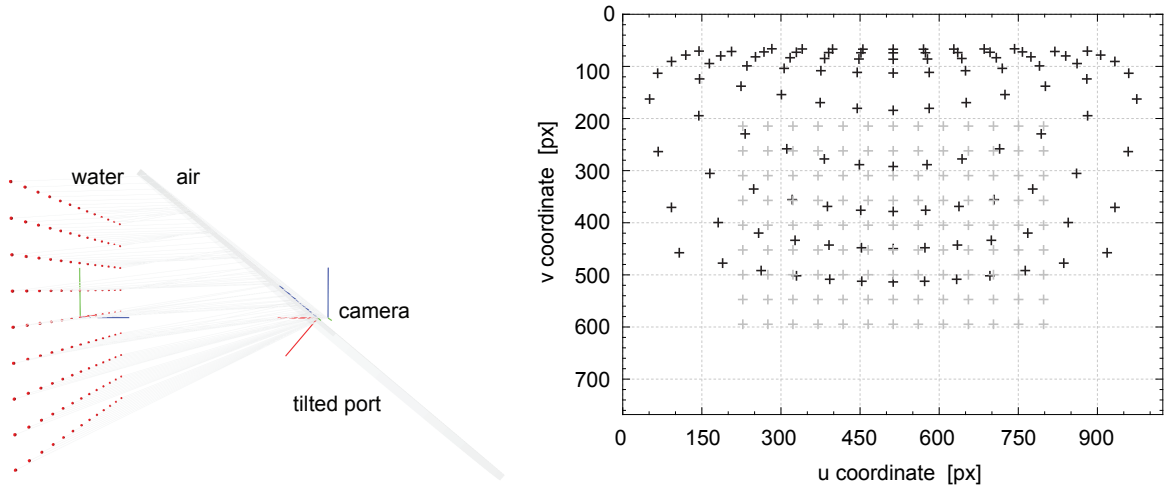

Figure 4.3.: Image of a point grid through a strongly tilted, large port (black crosses) in comparison to a perspective undistorted perception (grey crosses): The non-parallel port results in significant non-point symmetric image distortion (50 ° port pitch, 0.5 m point grid distance)

for a curve. In the upper and lower left section of the image (curves 1 and 7), the curves converge against a straight, outwards directed line. As the horizontal coordinates of the left image points decrease, the origins of their refractive epipolar curves (1, 2, 3, 7, 8 and 9) move towards the upper or lower image boundary. The visualisations show that refractive epipolar curves significantly differ from their perspective counterparts.

The camera image distortion caused by the refraction at a strongly tilted port is visualised in Fig. 4.3 using a front parallel flat grid of points. The image distortion shown significantly differs from the pincushion distortion of a parallel port depicted in Fig. 4.1. The distortion is neither point-symmetric nor correctable by standard decentering or tangential distortion models, respectively [59].

In Treibitz et al. [68, 71] the theoretical single viewpoint (SVP) model was described in which all rays meet in one common point, namely the camera's centre of projection $\boldsymbol{p}_c$, if the flat interface is exactly located on $\boldsymbol{p}_c$. In this case the distortion is correctable by a function, which depends only on the radial image coordinate and is object distance independent. We show in Fig. 4.4 that the SVP model only exists for small angles. Since the incident angle in water can physically not exceed the maximum angle of total internal reflection of $\arcsin n_a/n_w \approx 42\,^\circ$, a shift of the refrax $\boldsymbol{p}_{aw}$ occurs for larger angles, which results in a virtual focal point $\boldsymbol{p}$ that is different from the common centre of projection $\boldsymbol{p}_c$. Even though the SVP model is not defined for large angles, the physics-based distortion correction function of Treibitz et al. [68, 71] is still valid.

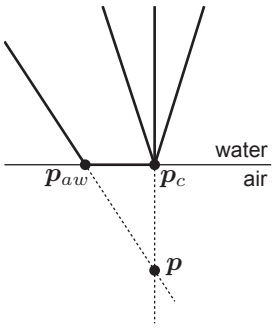

Figure 4.4.: Restriction of the validity of the SVP model: The image shows that the single viewpoint model is only applicable to small angles so that according rays meet in the common point $\boldsymbol{p}_c$. Larger angles result in a shift of the refrax $\boldsymbol{p}_{aw}$ with a different virtual focal point $\boldsymbol{p}$.

4.4. Proposed refractive magnification function for the analysis of refraction

In this section, we present a novel fundamental mathematical function that completely characterises the refractive magnification in the image of a camera with two refractive interfaces being parallel to the camera's image plane. The equation also incorporates the single interface case and can be easily extended for arbitrary multi interface systems. In contrast to previous literature in which the magnification effects by arbitrary thick ports were only partially described, this proposed function allows the comprehensive understanding of refractive magnification, its dependency to all parameters in the system and its exact calculation. The function also allows a simplified qualitative evaluation of refractive effects, and is an alternative to an exact quantitative calculation requiring a more complicated refractive forward projection, which is not available in a closed form.

The proposed function corresponds to an extension for arbitrary thick ports of the radial distortion correction function for single interfaces presented in [68, 71]. In contrast to a representation of the relation between the undistorted and distorted radial coordinates in the form of a radial distortion function, we prefer the representation of the occurring magnification.

The importance of our novel fundamental magnification function is the ability to describe and calculate numerous properties of an underwater camera with a thick, flat port being parallel to the camera's image plane (see section 4.4.1, 4.4.3, and 4.4.2).

4.4.1. Newly discovered properties and methods

In addition to known properties, the function also served the authors as a basis to present newly discovered properties and to obtain more accurate calculations of the optical system, such as:

1. Closed-form calculation of the magnifications occurring in the image perceived through a thick, flat and parallel port. The extension for multiple thick layers is straight forward.
2. Increase/decrease in magnification with increasing object distance depending on the port thickness and the port distance
3. The magnification of close objects at the image centre does not only depend on the refractive indices of air and water, but also on the object's distance and the port's distance, thickness and index of refraction.
4. Calculation of an optimal port distance for reduced object distance dependency
5. Increased object distance dependency with increasing distance to the characteristic radial distance
6. Closer objects are less distorted than more distant objects.
7. The magnification increases with growing port thickness.
8. Chromatic aberration is object distance dependent at small object distances and loses its distance dependency at large object distances.

The derivation of these properties and methods is described in more detail in section 4.5.

4.4.2. Proposed extensions of the radial distortion correction function for arbitrary thick ports

Treibitz et al. [68, 71] described the distance dependent radial distortion correction function for thin ports. Using our refractive magnification function, we extended their function for arbitrary thick ports and obtained the following results:

1. Precise radial distortion correction function for arbitrary thick ports, which are parallel to the image plane (see (4.7))
2. At large object distances the image magnification is identical to the thin flat interface case, which allows a simple correction of image distortion and a straight forward integration into the perspective camera model (see (4.8)).

4.4.3. Existing properties and methods

To date, different effects and properties of flat refractive geometry have been described. The following list summarises the effects and properties, which can be derived from our proposed fundamental refractive magnification function. The detailed derivations are omitted here.

1. In Gaussian optics, objects appear larger as if they were taken with a larger focal length, which is about n_w/n_a times larger than the actual focal length [63].
2. Distortion/magnification increases with radial distance (pincushion distortion) [58]
3. Increase in magnification with increasing object distance or decreasing port distance [68, 71]
4. Port distance independence if water is replaced by air [73]. The calibration of the port distance is not possible in air.
5. Object distance independent magnification with a thin port located on the camera's centre of projection [68, 71]
6. The magnification of objects at large distances is independent of the port's thickness [58], distance [68, 71] and index of refraction.
7. The change in the magnitude of magnification with respect to the port thickness is relatively small compared to the change caused by the radial distance [68, 71]
8. The magnification increases if the index of refraction of water or the index of refraction of the port increase. It decreases with increasing index of refraction of air.
9. As a result of the wavelength specific image magnification (chromatic aberration), red coloured fringes appear relative to the image centre on the inside and blue coloured fringes on the outside of edges in the image [63, 58].

4.4.4. Derivation of the refractive magnification function

In this section, the proposed refractive magnification function is derived. The magnification function describes the ratio between the radial distances r' and r of the image points of an object point at the distance z_p perceived by a perspective camera with and without a parallel port (see Fig. 4.5)

$$f_m(r', z_p) = r'/r. \tag{4.1}$$

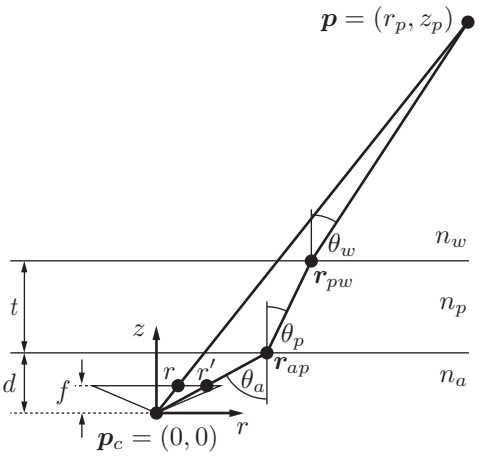

Figure 4.5.: Illustration to calculate the refractive magnification function

f_m is composed of a refractive back projection of the image point r' on an object plane, which is parallel to the camera's image plane and located at the object distance z_p, and a subsequent perspective forward projection. As the complicated refractive forward projection is not required, f_m can be solved in a less complex, closed form.

The refractive back-projection is calculated as follows. It starts with a perspective back projection of the image point r' onto the air-port interface. For the refrax on the air-port interface, we have

$$r_{ap} = \frac{r'd}{f}, \tag{4.2}$$

where f denotes the focal length of the camera and d the port distance (see Fig. 4.5). Secondly, the relation between the object point $\boldsymbol{p} = (r_p, z_p)$ and the refrax r_{ap} is defined by

$$r_p = r_{ap} + t \tan\theta_p + (z_p - t - d) \tan\theta_w, \tag{4.3}$$

with t denoting the thickness of the port, and θ_p and θ_w representing the refracted angle in the port and the incident angle in water. Both angles are obtained from the refracted angle θ_a in air using Snells law

$$n_a \sin\theta_a = n_p \sin\theta_p = n_w \sin\theta_w, \tag{4.4}$$

with n_a, n_p, n_w denoting the indices of refraction of air, the port and water. Here, $\sin\theta_a$

can also be described by

$$\sin\theta_a = \frac{r'}{\sqrt{r'^2 + f^2}}. \tag{4.5}$$

Lastly, the radial image coordinate of the object point $\boldsymbol{p}$ in the undistorted image corresponds to the perspective forward projection

$$r = \frac{r_p f}{z_p}. \tag{4.6}$$

Using (4.1), (4.2), (4.3), (4.4), (4.5) and (4.6), the proposed fundamental image magnification function yields

$$\begin{aligned} f_m(r', z_p) &= \frac{z_p}{(b_p - b_w)t + (1 - b_w)d + b_w z_p}, \\ b_i &= \frac{n_a f}{\sqrt{n_i^2 f^2 + (n_i^2 - n_a^2)r'^2}},\ i \in \{p, w\} \end{aligned} \tag{4.7}$$

Equation (4.7) can be obtained by a symbolic mathematics software or derived by hand. The derivation by hand is described in more detail in section A.1. Note that the domain of b_i is $0 < b_i \leq \frac{n_a}{n_i}$. Also note, that an extensions for more than one thick layer is easily possible by adding further addends $(b_i - b_w)t_i$ to the denominator of $f_m(r', z_p)$, similar to $(b_p - b_w)t$ for the port layer.

4.5. Newly observed and extended characteristics of refractive magnification of a thick port

We would like to start the description of the novel observed characteristics of refractive magnification with three plots (see Fig. 4.6). The plots of refractive magnification differ in the port thicknesses (1.2 cm, 6 cm and 4 cm) and port distances (2 cm, 1 cm and 1 cm) and are plotted as a function of the radial distance r' and the object distance z_p.

In all three examples, the magnification $f_m(r_r)$ has its minimum in the image centre and grows with r'. The magnification converges for large object distances against a function, which is independent from z_p. From (4.7) follows

$$\lim_{z_p \to \infty} f_m(r') = \frac{\sqrt{n_w^2 f^2 + (n_w^2 - n_a^2)r'^2}}{f n_a}, \tag{4.8}$$

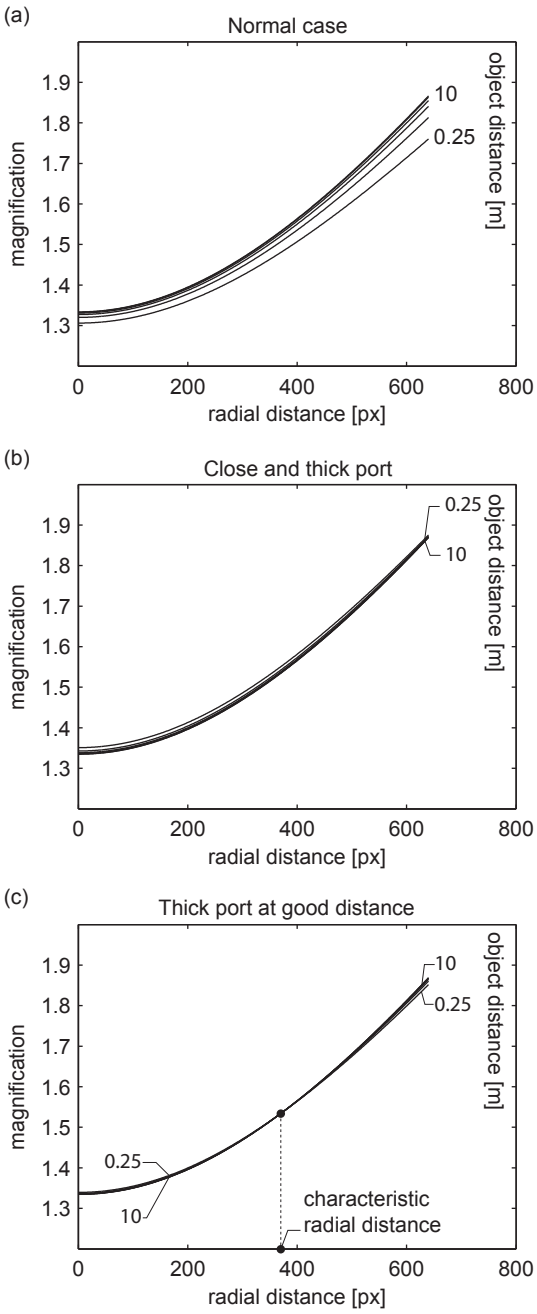

Figure 4.6.: Refractive magnification and its object distance dependency: The magnification r'/r is plotted as a function of the radial distance r' in the distorted image and as a function of different object distances (0.25, 0.5, 1, 2, 5 and 10 m). Figure (a) shows a common case (port thickness $t = 12\,\text{mm}$, port distance $d = 2\,\text{cm}$) in which the magnification grows with object distance. Figure (b) depicts the magnification of a thick, close port ($t = 6\,\text{cm}$, $d = 1\,\text{cm}$) for which the magnification decreases with increasing object distance. The last case ($t = 4\,\text{cm}$, $d = 1\,\text{cm}$) shows an example of a simultaneous increase and decrease in image magnification and a reduced distance dependency. The characteristic radial distance is highlighted.

which is completely independent of the port's distance, thickness and refractive index. In accordance with Wakimoto [58] and Treibitz et al. [68, 71], the only essential property of the port at large distances is its orientation, and therefore, (4.8) is equivalent to the single interface case.

Equation (4.8) shows that the radial distortion for objects located at a sufficiently large distance can be well approximated by a constant function without the necessity to consider the scene depth. Moreover, as the distortion is known, the standard perspective camera model with a modified focal length in combination with a physics based image distortion correction as proposed by Treibitz et al. [68, 71] for the single interface case can be used.

Next, we would like to describe a novel finding describing the object distance dependency of the image magnification, which only occurs for thick ports. The image magnification does not necessarily increase with the port distance as described in [71] for the single interface case and as shown in Fig. 4.6a for thick ports. It can also decrease with increasing scene depth (Fig. 4.6b) or can show increasing and decreasing magnification effects at the same time (Fig. 4.6c). The change in image magnification depends on the port thickness and the port distance, as will be shown next.

As is well known, refractive image magnification grows with radial distance. As shown in Fig. 4.7, the radial pixel shift in the camera image increases with increasing distance between the imaged object point and the optical axis ($\boldsymbol{p}_1 \rightarrow \boldsymbol{p}_3$). Pincushion distortion occurs, as the radial pixel shift caused by refraction increases more quickly than the radial pixel shift of a perspective projection. Considering the extended rays in water and their intersection with the optical axis (Fig. 4.7), the virtual focal points do not coincide at one point but are shifted on the optical axis against the viewing direction of the camera with increasing object point to optical axis distance. A constant magnification with respect to the perspective camera model, however, requires the ray in water of an object point to be located on its perspective projection line. In this case, the ray in water and the projection line constitute a common focal point or a common centre of projection, respectively.

If the focal point of the ray in water is located behind the camera's centre of projection and the object point is moved on its projection line away from the camera (see Fig. 4.7, $\boldsymbol{p}_1 \rightarrow \boldsymbol{p}_2$), the image point of the refracted light ray in air is, similarly to the $\boldsymbol{p}_1 \rightarrow \boldsymbol{p}_3$ case, shifted away from the optical axis resulting in an increased magnification of the object in the camera image.

The opposite effect, which has not been described in literature yet, occurs if the port

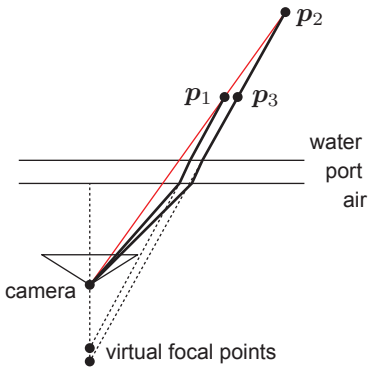

Figure 4.7.: Image magnification - normal case: (1) Refraction increases the incident angle of light rays of a perspective camera (red: perspective, black: refracted light ray path). (2) Pincushion effect: The magnification effect increases with the radial distance of the object point ($\boldsymbol{p_1} \rightarrow \boldsymbol{p_3}$) and results in distinct virtual focal points. (3) The magnification also depends on the object point's distance to the camera. If the virtual focal point of the ray in water is located behind the camera, the magnification increases with the object point distance ($\boldsymbol{p_1} \rightarrow \boldsymbol{p}_2$), else it decreases (see Fig. 4.8).

is relatively thick with respect to its distance to the camera. In this case, the focal points of the extended rays in water are located in front of the camera (see Fig. 4.8). Increasing the distance of the object point to the camera along its perspective projection line ($\boldsymbol{p}_1 \rightarrow \boldsymbol{p}_2$) results in a shift of its image point towards the optical axis and therefore in decreased image magnification. There is also a third case, in which the increasing and decreasing magnification effect can be observed simultaneously. Due to the axial camera properties of the underwater camera [73] and as depicted in Fig. 4.7 and Fig. 4.8, the virtual focal points of the extended rays in water do not necessarily coincide in one common point. It is possible that the virtual focal points are distributed on the optical axis around the cameras centre of projection. In this case, the growth in object distance is connected with an increase in magnification of image regions greater than a certain radial distance and a simultaneous decrease in magnification in the remaining circular image region around the image centre. We call this radial distance characteristic radial distance. It is highlighted in Fig. 4.6c.

The configuration from which the magnification decreases with growing object distance can be computed by the magnification function around the image centre:

$$f_m(r')\Big|_{r'=0} = \frac{n_{pw} z_p}{(n_{aw} - n_{ap})t + (n_{pw} - n_{ap})d + n_{ap} z_p}, \tag{4.9}$$

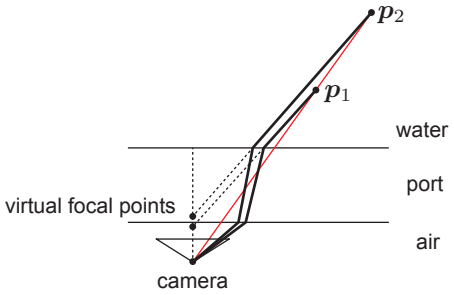

Figure 4.8.: Image magnification - thick close port case: If the port thickness-to-camera port distance ratio is high, the virtual focal points are located in front of the actual focal point of the camera. In this case, the magnification decreases with growing object point distance.

where $n_{ij} = n_i n_j$. The term $k = (n_{aw} - n_{ap})t + (n_{pw} - n_{ap})d$ of the denominator allows the direct calculation of the point where the magnification of the central image regions starts decreasing. If $k = 0$, f_m is constant and the virtual focal points for angles equal zero coincide with the centre of projection of the camera as discussed above and form a perspective system. In this case, the magnification at the image centre $\boldsymbol{u}_0$ is constant, that is $n_w/n_a \approx 4/3$, and the distance dependency of the image region around $\boldsymbol{u}_0$ is small. Considering the work in [68] or [71], respectively, this case is similar to their presented SVP approximation. The tip of the caustic is directly located on the cameras centre of projection, and a perspective model with an adapted lens distortion model can be used to approximate image distortion for small angles. If $k < 0$, the novel, described effect occurs in which the magnification decreases within a certain circular image region around $\boldsymbol{u}_0$ for growing object distances. For $k < 0$, the following relation between the port distance and the port thickness holds:

$$d < \frac{n_{aw} - n_{ap}}{n_{ap} - n_{pw}} t, \tag{4.10}$$

meaning in our setup that the distance of the port needs to be about 0.32 times smaller than its thickness.

Equation (4.9) also shows that the magnification of close objects at the image centre does not only depend on the refractive indices of air and water, but also on the object's distance and the port's distance, thickness and index of refraction. This stands in contrast to [68, 71].

A general equation to describe the optimal port distance d to achieve perspective

behaviour at the characterisitc radial distance r_c is given next:

$$d = \frac{f n_a (c_p - c_w) t}{c_p (c_w - f n_a)},$$
$$c_i = \sqrt{n_i^2 f^2 + (n_i^2 - n_a^2) r_c^2},\ i \in \{p, w\}. \tag{4.11}$$

The equation was obtained by setting the term $(b_p - b_w)t + (1 - b_w)d$ of (4.7) equal zero. As shown in Fig. 4.6c and in contrast to Fig. 4.6a, the distance dependency can be significantly reduced, if r_c lies in the middle of the visible radial distance range. The distance dependency gets more distinct the more r_c differs from this point. In Fig. 4.6c, the characteristic radial distance is located at about the middle of the available radial distance domain at $r_c = 371\,\text{px}$, in Fig. 4.6b at about 718 px (not shown) and in Fig. 4.6a at $-\infty$.

The optimal port distance, however, is relatively small and can be only achieved for relatively thick ports and lenses with a short focal length. In our setup with a very short focal length of 2 mm and a relatively thin port, the optimal distance between the port and the optical centre of the camera results in about 3.4 mm, with r_c set to a quarter of the image width. Incorporating the additional spacing by the housing of the stereo camera itself, the optimal distance is practically not realisable, and therefore, the camera should be placed as close as possible to the port to reduce the object distance dependency.

As especially visible in Fig. 4.6a, the object distance dependency also grows with increasing radial distance to the characteristic radial distance r_c. To simplify the equations and to support the understanding of the characteristics of the magnification at large radial distances, we analyse the convergence behaviour of the magnification for large radial distances. For large radial distances, the refracted angle θ_a in air is nearly $\pi/2$ and the refracted angle in the port θ_p converges against its critical angle of total internal reflection. Thus, the total internal reflection in the port determines the maximum angle θ_w in water. A change of the radial distance result only in a shift of the light rays in the port and in water along the interfaces. Using (4.7) and presuming a close object, the resulting equation shows that the magnification converges against a constant value, which only depends on the object distance-to-port distance ratio

$$\lim_{r' \to \infty} f_m(r') = \frac{z_p}{d}. \tag{4.12}$$

The port thickness and indices of refraction have less influence on the refraction be-

haviour with growing radial distance. For large radial distances at constant small object distances, the system becomes perspective. A flat object, which is parallel to the port and observed at a large angle, appears mainly magnified and little distorted. In contrast to that, $f_m(r')$ converges against a linear function of r' for large object distances

$$\lim_{p_z \to \infty} f_m(r')\Big|_{r' \gg f} = \frac{\sqrt{n_w^2 - n_a^2}}{n_a f} r'. \tag{4.13}$$

An object appears strongly distorted here. The convergence behaviour of $f_m(r')$ against a constant value or a linear function for small or large object distances, respectively, demonstrates the growing distance dependency of $f_m(r')$ with increasing radial distance. Whereas at small radial distances the difference in magnification between close and distant objects is relatively small, it increases with growing radial distance.

Closer objects are less distorted than more distant objects. It is known that $f_m(r')$ represent a pincushion distortion and is, thus, a monotonically increasing function. As a result, the range of $f_m(r')$ is limited to

$$f_m(r')\Big|_{r'=0} \leq f_m(r') \leq \lim_{r' \to \infty} f_m(r') = \frac{z_p}{d}. \tag{4.14}$$

The width $w = z_p/d - f_m(r')\Big|_{r'=0}$ of this range decreases with decreasing object distance z_p, as

$$\frac{\partial w}{\partial z_p} = \frac{1}{d} + \frac{n_p n_w (d n_p (n_a - n_w) + n_a (n_p - n_w) t)}{(d n_p (n_a - n_w) - n_a (n_p (z_p - t) + n_w t))^2} > 0. \tag{4.15}$$

For proofing the correctness of (4.15), it is obvious that the first addend of (4.15) and the denominator of the second addend are positive. We can also show that the enumerator of the second addend is positive and, thus, (4.15) is correct. Therefore, we can write

$$n_p n_w (n_p (n_a - n_w) d + n_a (n_p - n_w) t) > 0 \tag{4.16}$$

$$\Leftrightarrow \frac{n_p (n_w - n_a) d}{n_a (n_p - n_w) t} > 1, \tag{4.17}$$

which is true as $n_p > n_a$, $n_w - n_a > n_p - n_w$, and $d > t$. Since w increases with z_p, the magnification $f_m(r')$ changes less with respect to r' for close object distances. Hence, closer objects are less distorted than more distant objects.

The effect of the port thickness t has only been briefly described in literature [68, 71].

The direction, in which f_m changes depending on t, was not mentioned. According to (4.7), the magnification f_m increases with growing port thickness t

$$\uparrow t \Rightarrow \uparrow f_m(r'). \tag{4.18}$$

Even though the proportion on the overall magnification is low, we show in section section 4.7 that the port thickness still needs to be modelled for close range applications.

The magnification function also allows proofing the distance dependency of lateral chromatic aberration (LCA) for small object distances and the proof of its distance independence for large object distances. LCA occurs due to the wavelength dependency of the refractive indices (dispersion) resulting in wavelength specific magnification. Considering two different wavelengths with corresponding different indices of refraction in water, the port and air, $n_{a,i}, n_{p,i}, n_{w,i}$, we have for the LCA

$$\begin{aligned} f_{ca}(r', z_p, \boldsymbol{n}) &= |f_m(r', z_p, \boldsymbol{n}_1) - f_m(r', z_p, \boldsymbol{n}_2)|, \\ \boldsymbol{n} = (\boldsymbol{n}_1^T, \boldsymbol{n}_2^T)^T,\ \boldsymbol{n}_i &= (n_{a,i}, n_{p,i}, n_{w,i})^T,\ \boldsymbol{n}_1 \neq \boldsymbol{n}_2 \end{aligned} \tag{4.19}$$

The LCA changes with the object distance z_p, and as a result of (4.8), f_{ca} converges at large z_p against a function, which is independent from z_p. A quantitative description of the extent of LCA is given in section 4.6.

4.6. Proposed quantative analysis of chromatic aberration

Chromatic aberration (CA) is a well-known phenomenon of flat port underwater cameras. CA was especially described in scientific literature of the 1960s [63, 74, 58] and by recent underwater photography literature, but is hardly mentioned in recent underwater photogrammetry publications. Whereas Sedlazeck and Koch [75] considered the variation of the refractive index as minor and neglected it, Yau et al. [72] are one of the few groups explicitly exploiting CA. In this chapter, we present the extent of CA in a wide-angle underwater camera system as a function of radial distance and show how CA changes with object distance.

Chromatic aberration is caused by wavelength dependent refraction of light, called dispersion. The propagation direction of light with a longer wavelength is less refracted than light with a shorter wavelength [76]. Chromatic aberration is subdivided into

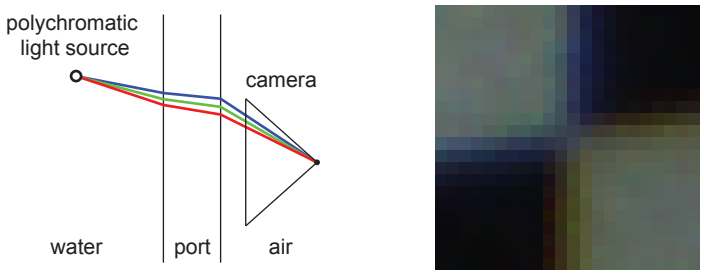

Figure 4.9.: Lateral chromatic aberration: Light with a longer wavelength is less refracted than light with a shorter wavelength (left) becoming visible in the form of colour fringes as shown for the lower right image section (right)

axial and lateral chromatic aberration. Axial chromatic aberration (ACA) describes the wavelength dependent focus of light beams. Depending on the wavelength, the beams are focused at different distances to the camera lens. Since we use a pinhole camera model in which light beams are always focused, ACA does not occur in our analysis. In contrast to that, LCA is visible in flat refractive geometry. Light of shorter wavelength (e.g. blue) is projected at a larger radial distance than light of longer wavelength (e.g. red) [58, 1]. In Fig. 4.9, LCA is visible as colour fringes at sharp edges (Fig. 4.9b) and is especially visible at large radial distances in the image corners [58].

The effect of chromatic aberration is particularly observable in our underwater camera setup with a flat port, a short focal length and a relatively high resolution. To measure the extent of LCA numerically, the physics-based refractive camera model was used. Using the calibration data of our stereo camera with a displaced centre of projection and a port assumed to be ideally parallel with the cameras' image planes, the maximum radial distance amounts to about $\hat{r}' = 653\,\text{px}$. LCA is measured as the difference between the blue and red image points representing the wavelengths λ_b (463 nm) and λ_r (630 nm). Whereas a qualitative analysis of LCA, as used in (4.19), can be based on straight forward refractive back projection (section 5.3.2), an accurate quantitative calculation of LCA requires more complex refractive forward projection (chapter 5).

Fig. 4.10 depicts the magnitude of LCA of an object point at a distance of 1 m as a function of the radial distance r'. The curve exhibits an almost constant slope for small angles, rises more quickly with larger angles and reaches a considerable maximum of 9.7 px at $\hat{r}'$, which is clearly visible and shouldn't be neglected.

The distance dependency of LCA at $\hat{r}'$ is depicted in Fig. 4.11 for close distances of 0.2 to 1 m together with the maximum LCA at infinity. The curve confirms the distance dependency of LCA for close distances and the distance independence for large object

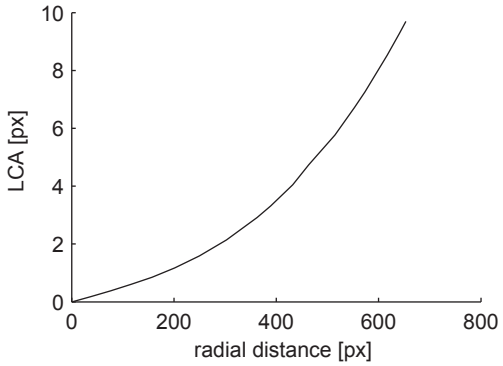

Figure 4.10.: LCA as a function of the radial distance: The graph shows the magnitude of LCA between the blue and red projections with an object distance of 1 m.

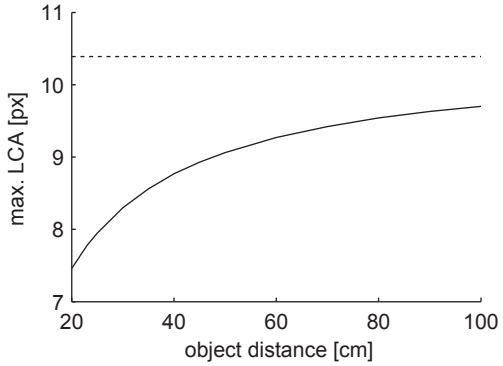

Figure 4.11.: LCA as a function of the object point distance (solid line) at maximum radial distance. The curve is shown for short distances and converges with increasing object distances against its maximum (dotted line).

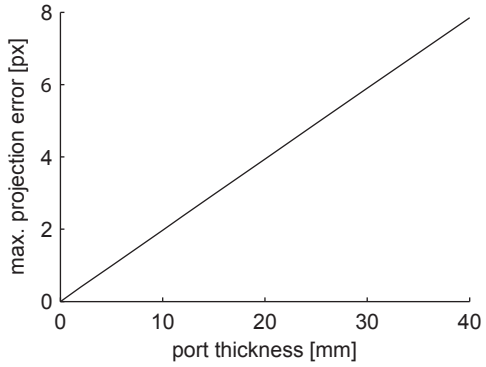

Figure 4.12.: The image error caused by assuming a thin instead of a thick port: The image error of a 1 m distant object point perceived at maximum radial distance grows almost linearly with 2 px/cm.

distances as described in section 4.6. LCA increases with distance and converges for points at infinity against a constant value (10.4 px). The difference in LCA between points at 0.2 (7.5 px) and 1 m (9.70 px) amounts to 2.2 px.

Note that the resolution of our 0.8 megapixel (MP) photogrammetric video camera is relatively low compared to single shot photogrammetric cameras or high resolution consumer grade video cameras. The LCA in pixels can easily increase by one order of magnitude for those high resolution image sensors.

4.7. Proposed quantitative analysis of the refractive effect caused by the thickness of the port

To measure the refractive effect caused by the thickness of a port, the two image points of an object point, perceived through a thin and thick port with thickness t and distance $d = 2\,\text{cm}$ were compared. The object point was located at a distance of $z_p = 1\,\text{m}$ and appeared in the camera image with a thick port at maximum radial distance. The physics-based refractive camera model was used again to compute the projection error $e_t(t)$, which occurs if a thin instead of a thick port is assumed. $e_t(t)$ is depicted in Fig. 4.12 as a function of the port thickness $t \in [0, 4\,\text{cm}]$. Fig. 4.12 shows that the projection error grows almost linearly with the port thickness in the value range considered with a slope of about 2 px/cm.

The analysis of the refractive effect of the port thickness shows that the image error is relatively small in comparison to the effect of chromatic aberration [68], but, in the

opinion of the authors, still significantly high for accurate image restoration and 3D reconstruction. Therefore, the port thickness needs to be integrated into the camera model.

4.8. Conclusion

In this chapter, the effects of refraction on image perception with one or two flat refractive interfaces were evaluated. At first, as a clarification of prior research, this chapter visualised the refractive epipolar curves in the whole camera image, visualised the image distortion with a non-parallel port, and revealed the limitation of the SVP model. It was shown, that the curvature of the refractive epipolar curve is particularly strong in the upper right and lower right region of the right camera image and exhibits a maximum change in the vertical direction of 30 px using our underwater stereo camera. For non-parallel ports the image distortion was visualised, which significantly differs from the pincushion distortion with parallel ports, and is obviously neither point-symmetric nor correctable by standard decentering or tangential distortion models. Furthermore, it was shown that the SVP model is only valid up to a maximum incident angle in water of about 42°.

In the major part of this chapter, the fundamental refractive magnification function was proposed. This function enabled the description of at least 10 new and 9 known effects or methods for flat port underwater cameras. In addition to that, the extent of lateral chromatic aberration and the effect of the thickness of the port was analysed for the wide-angle underwater camera used. The analysis revealed a high maximum lateral chromatic aberration of about 10 px and a projection error of about 2 px/cm with respect to the thickness of the port.

5. Real-time, memory efficient, accurate refractive forward projection

5.1. Introduction

Efficient refractive forward projection is essential in real-time applications, which use fast refractive calibration, bundle adjustment, simultaneous localisation and mapping (SLAM), structure from motion or image restoration. But Jordt-Sedlazeck et al. [78, 79, 77, 80, 75] and Kang et al. [81] hold the opinion that refractive forward projection is computationally too costly for calibrating underwater cameras with a thick flat port or for structure from motion in real-time. This view is incorrect, as we will show in this chapter.

Our research in this chapter is motivated by our need for real-time, memory efficient, accurate refractive forward projection in underwater stereo cameras with a flat port. There are publications in which the port of an underwater housing or the wall of a fish tank are poorly approximated by a single interface to reduce the computational complexity of refractive forward projection through more than one interface [82, 68, 71]. With such approaches, a simplified forward projection can be realised by just solving the roots of a quartic equation in a closed form [83]. To this day however, almost no publications exist in which the execution times of various refractive forward projection methods were actually compared. Kawahara et al. [2] measured the execution time of their own refractive forward projection method and compared it with their implementation of the standard back projection-based refractive forward projection method and with Agrawal et al.'s [73] method, but did not consider the accuracy of the methods. Furthermore, all methods were implemented and measured in Matlab, which does not allow a direct time comparison with implementations in C++, which is often used for efficient image processing. In contrast to prior literature, we will give an overview on

the efficiency of various refractive forward projection methods. For fair comparison, the fastest forward projection methods are benchmarked under the same accuracy criterion.

In this section, we firstly evaluate the memory requirements of Maas's [1] 2D and Kawahara et al.'s [2] 1D lookup table based methods to guarantee minimum radial refractive back projection accuracy up to a certain object distance. Moreover, we propose a more efficient forward projection method for applications where lookup table-based methods are too time-consuming or require to much memory. Our proposed method improves applications such as the calibration of underwater cameras, the forward projection with changing indices of refraction or the forward projection with light consisting of multiple wavelengths. Additionally, we propose an efficient lookup table-based forward projection method, which is slightly slower than Maas's method, but requires less memory. To complete the list of refractive forward projection methods, we also discuss polynomial equation-based refractive forward projection methods and their limitations. Finally, we compare various refractive forward projection methods by measuring their execution times.

5.2. Related literature

In 1988, Kotowski [84] presented a forward projection method for refractive systems consisting of multiple interfaces with various shapes. This proposed solution of a non-linear equation system was based on numerical minimisation using Newton's method. Due to the general interface shape model, the minimisation took place in the three-dimensional space of the refrax, which is computationally expensive and prone to getting stuck in local minima. Kunz and Singh [59] also used numeric optimisation to obtain the refrax on flat or dome ports. However, they provided no detailed description of their method.

In 1995, Maas [1] also used numerical minimisation-based refractive forward projection to compare its performance with the actually proposed lookup table-based refractive forward projection method (see below). In contrast to Kotowski, the numerical minimisation took place in the plane of incidence with a one-dimensional search space, and, hence, is more efficient and stable. Moreover, the method of Maas handled the refraction at two interfaces. In 2013, Kawahara et al. provided a complete equation set for the implementation with Newton's method. The equations were inefficient, because they required the computation of many time-consuming reciprocal square roots and used an initial guess based on perspective projection. We will propose an alternative Newton

method-based approach, which is significantly faster than Kawahara et al.'s method and reduces the execution time by more than 43 % from 304 ns to 171 ns.

Since numerical minimisation-based refractive forward projection with two interfaces was too time-consuming, Maas [1] proposed in 1995 a 2D lookup table in combination with interpolation. The input parameters of the lookup table were defined by the indices of discrete coordinates of the plane of incidence. This lookup table-based forward projection method has the advantage of being faster and is independent of the number of interfaces. Since Maas' lookup table works only in the plane of incidence, the execution time of this proposed method is mainly determined by the transformation time from object coordinates to plane of incidence coordinates, the transformation back, and the perspective camera projection. A general disadvantage of such lookup table-based approaches is the need to initialise the lookup table. For this reason, Maas' method is not applicable to the port pose calibration, for example. Unfortunately, Maas neither described the lookup table size required for a certain accuracy and distance nor reported on the execution time of his method. In this section, we evaluate the necessary lookup table size for a back projection accuracy of 0.01 mm for a distance of up to 1 m and measure the execution time required. Our results show that Maas' method is significantly faster than all other methods benchmarked. We also show that the main disadvantage of Maas' method lies in the huge memory requirement for large visual ranges and high accuracies. The disadvantage of this and all other lookup table-based methods presented in this chapter is that the memory required multiplies with the number of wavelengths and cameras used.

In 2000, Glaeser and Schröcker [83] analytically deduced the position of the refrax for the refraction at one flat interface. Based on Fermat's principle, the light ray with the least travel time between the object point in water and the camera in air was defined. The equation for the radial coordinate of the refrax represents a polynomial of degree four, which is solvable in a closed form. The advantage of this method is based on its high accuracy. However, Glaeser and Schröcker's theoretical work does not elaborate on the practical processing time required to solve the quartic equation. In this chapter, we measure the execution time of Glaeser and Schröcker's solution to the forward projection problem in combination with the analytical quartic equation solver of Herbison-Evans [85].

In a recent paper of 2012, Agrawal et al. [73] derived an alternative solution to the forward projection problem with two parallel flat interfaces. This proposed approach is based on finding roots of a 12th degree polynomial. Since closed-form solutions are only

known for polynomials up to a degree of four, Agrawal et al.'s method needs iteration-based polynomial root finders. In comparison to Glaeser and Schröcker's [83] quartic equation, where the correct root lies in a unique interval, the selection of Agrawal et al.'s correct root from the computed real roots requires repetitive computations of back projection errors until the corresponding root is found. Kawahara et al. [2] measured the execution time of Agrawal et al.'s method and demonstrated its inferiority in efficiency. Our more comprehensive benchmark reveals that Agrawal et al.'s method is even two orders of magnitude slower than the most efficient refractive forward projection algorithm of Maas [1]. Our analysis also revealed another disadvantage to Agrawal et al.'s approach, where polynomial root solvers were only able to find correct roots up to a limited object distance. In our scenario, all root solvers were unable to calculated the correct solutions for object distances greater than 10 m.

Jordt-Sedlazeck and Koch claimed that direct refractive forward projection is extremely time-consuming for the calibration of underwater cameras [75] or infeasible for large non-linear optimisation problems like bundle adjustment [79]. In contrast to that, we will show that refractive forward projection can be executed efficiently enough for real-time processing.

The recently presented method of Kawahara et al. [2] calculates refractive forward projection by a recurrence relation and a one-dimensional lookup table. In terms of memory requirement, Kawahara et al.'s lookup table-based method is superior to Maas's [1] approach. Compared to the 2D lookup table-based method, the size of Kawahara et al.'s 1D lookup table does not grow with the maximum object distance. It also only increases linearly with the accuracy required. The execution time of Kawahara et al.'s method is also independent of the number of interfaces. Like for all lookup table-based methods, this algorithm is only applicable to calibrated underwater cameras with known constant indices of refraction. Another disadvantage of the method lies in its execution time. In comparison to Maas [1] 2D lookup table-based method, Kawahara et al.'s algorithm requires several iterations and is, thus, more inefficient. Kawahara et al. have compared the execution times of their 1D lookup table-based algorithm with a traditional back projection-based forward projection method. However, in their comparison, the lookup table size, which is necessary to achieve certain accuracy, was not considered. In this chapter, we also propose the lookup table size that is necessary to guarantee minimum back projection accuracy up to a certain object distance.

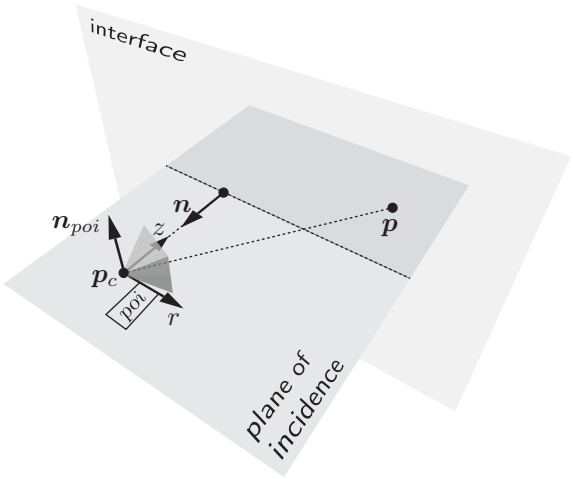

Figure 5.1.: Derivation of the POI coordinate system

5.3. Background

5.3.1. Coordinate transformations between 3D object space and 2D plane of incidence

Working in the 2D plane of incidence (POI) instead of the 3D object space leads to simplified equations and computationally more efficient and stable algorithms. The plane of incidence is particularly useful in polynomial equation-based refractive forward projection (see section 5.6) and numeric optimisation-based (see section 5.5.1) refractive forward projection.

As briefly described by Agrawal et al. [73], working in the plane of incidence requires the transformation from object space coordinates to plane of incidence coordinates

$$\begin{aligned} f_{poi} :\ & \mathbb{R}^3 \to \mathbb{R}^2, \\ & \boldsymbol{p} \mapsto (r, z)^T \end{aligned} \tag{5.1}$$

and the inverse mapping $f_{poi}^{-1} :\ \mathbb{R}^2 \to \mathbb{R}^3$ from plane of incidence coordinates back to object space coordinates. The plane of incidence coordinate system with its radial coordinate r and distance coordinate z is shown in Fig. 5.1.

As illustrated in Fig. 5.1, the plane of incidence is defined by the object point $\boldsymbol{p}$, the centre of projection of the camera $\boldsymbol{p}_c$ and the normal $\boldsymbol{n}$ of the interface given in object

space coordinates. The normal of the plane of incidence is calculated by

$$\boldsymbol{n}_{poi} = \frac{\boldsymbol{n} \times (\boldsymbol{p} - \boldsymbol{p}_c)}{|\boldsymbol{n} \times (\boldsymbol{p} - \boldsymbol{p}_c)|}, \tag{5.2}$$

or $\boldsymbol{n}_{poi}$ is an arbitrary unit vector perpendicular to $\boldsymbol{n}$, if $|\boldsymbol{n} \times (\boldsymbol{p} - \boldsymbol{p}_c)| = 0$. The plane of incidence coordinate system is centred at $\boldsymbol{p}_c$. Its orthogonal axes $\boldsymbol{r}$ and $\boldsymbol{z}$ lie in the plane of incidence and form the orthogonal matrix

$$\boldsymbol{M} = (\boldsymbol{r}, \boldsymbol{z}) = ((\boldsymbol{n}_{poi} \times \boldsymbol{n}), -\boldsymbol{n}). \tag{5.3}$$

$\boldsymbol{M}$ is used for the mapping from object coordinates to plane of incidence coordinates

$$(r, z)^T = \boldsymbol{M}^T(\boldsymbol{p} - \boldsymbol{p}_c), \tag{5.4}$$

and for the inverse transformation from plane of incidence coordinates to object space coordinates

$$\boldsymbol{p} = \boldsymbol{M} \begin{pmatrix} r \\ z \end{pmatrix} + \boldsymbol{p}_c. \tag{5.5}$$

All refractive forward projection methods $f_{rf} : \mathbb{R}^2 \to \mathbb{R}$ discussed in the next section work in the plane of incidence. The complete 3D refractive forward projection $f_{rf} : \mathbb{R}^3 \to \mathbb{R}^2$ consists of 4 steps, namely a transformation f_{poi} from object to plane of incidence coordinates, a 2D refractive forward projection f_{rf}, a coordinate transformation f_{poi}^{-1} back to the 3-dimensional object space, and a subsequent perspective forward projection f_{pf}

$$\begin{aligned} f_{rf} : \mathbb{R}^3 &\to \mathbb{R}^2, \\ \boldsymbol{p} &\mapsto f_{pf}(f_{poi}^{-1}(f_{rf}(f_{poi}(\boldsymbol{p})))). \end{aligned} \tag{5.6}$$

The perspective forward projection f_{pf} is defined by

$$\begin{aligned} f_{pf} : \mathbb{R}^3 &\to \mathbb{R}^2, \\ \boldsymbol{p} &\mapsto \boldsymbol{u}, \\ {}^c\boldsymbol{p} &= \ominus \boldsymbol{\pi}_c \oplus \boldsymbol{p}, \\ \boldsymbol{u} &= \frac{f}{{}^cp_z} \begin{pmatrix} {}^cp_x \\ {}^cp_y \end{pmatrix} + \boldsymbol{u}_0, \end{aligned} \tag{5.7}$$

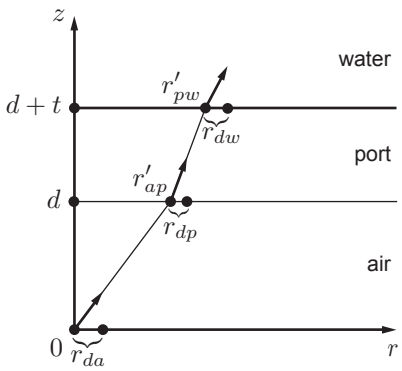

Figure 5.2.: Illustration of the parameters involved in the calculation of the refractive back projection in the plane of incidence

where $\boldsymbol{p}$ denotes the object point to be projected, $\boldsymbol{u}$ the resulting image point, $\boldsymbol{\pi}_c$ the pose of the camera, $\oplus$ the pose compounding operator, $\ominus$ the inverse pose operator (see section A.2), f the focal length and $\boldsymbol{u}_0$ the principle point of the camera.

In section 5.4, we evaluate prior approaches to the two-dimensional refractive forward projection problem $f_{rf} : \mathbb{R}^2 \to \mathbb{R}$, and then propose significantly improved solutions to f_{rf}.

5.3.2. Refractive back-projection in the plane of incidence

In this chapter, refractive back projection is used for our proposed more efficient back projection-based refractive forward projection method (see section 5.5.1) and to benchmark the evaluated refractive forward projection methods under fair conditions with the same back projection accuracy.

As shown by Kawahara et al. [2], performing refractive back-projection in the plane of incidence instead in the object space simplifies the refractive back projection equation significantly. In the plane of incidence, radial coordinates are sufficient to completely model refractive back projection, since the distances of the interfaces to the camera are known and the direction vectors of the rays are normed and point towards the object point. Hence, refractive back-projection in the plane of incidence can be described by

$$\begin{aligned} f_{rb} :\ & \mathbb{R} \to \mathbb{R} \times \mathbb{R}, \\ & r'_{ap} \mapsto (r'_{pw}, r_{dw}), \end{aligned} \tag{5.8}$$

where r'_{ap} denotes the refrax on the air-port interface and (r'_{pw}, r_{dw}) denote the origin

(refrax) and direction of the ray in water (see Fig. 5.2). To calculate f_{rb}, we start with the direction of the ray in air, which is defined by

$$r_{da} = \frac{r'_{ap}}{\sqrt{r'^2_{ap} + d^2}}, \tag{5.9}$$

where d denotes the distance of the air-port interface. Using Snell's law $n_i \sin\theta_i = n_j \sin\theta_j$, which simplifies to $n_i r_{di} = n_j r_{dj}$, as $r_{di} = \sin\theta_i$, the direction vectors in the port and water yield

$$r_{dp} = \frac{n_a}{n_p} r_{da},\ r_{dw} = \frac{n_a}{n_w} r_{da}. \tag{5.10}$$

Using the ray (r'_{ap}, r_{dp}) in the port to compute the refrax r'_{pw} on the port-water interface, we have

$$r'_{pw} = r'_{ap} + t \tan\theta_p, \tag{5.11}$$

where $\tan\theta_p = \sin\theta_p / \cos\theta_p$ simplifies in the plane of incidence to

$$\tan\theta_p = \frac{r_{dp}}{\sqrt{1 - r^2_{dp}}}. \tag{5.12}$$

A simplified representation of r'_{pw} is

$$r'_{pw} = r'_{ap} \left(1 + \frac{n_a t}{\sqrt{n_p^2 d^2 + (n_p^2 - n_a^2) r'^2_{ap}}} \right). \tag{5.13}$$

Thus, the refractive back-projection f_{rb} is fully defined and will be used for the back projection-based refractive forward projection and for benchmarking the forward projection methods under fair conditions with the same back-projection accuracy (see section 5.4.1, 5.4.2, 5.5.1, 5.5.2, and 5.7).

5.4. Proposed evaluation of the size of the lookup table of prior refractive forward projection methods

For comparing the lookup table sizes of different algorithms, we firstly evaluate the complexity classes of the lookup table size with respect to several criteria and subsequently

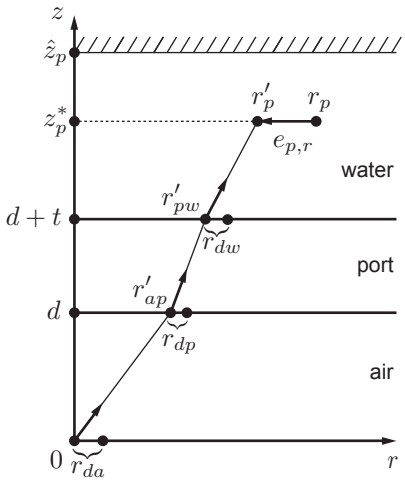

Figure 5.3.: Radial refractive back projection error $e_{p,r}$: $e_{p,r}$ is an efficient cost function in back projection-based refractive forward projection methods. In this chapter, $e_{p,r}$ is also used to define a common accuracy criterion in order to determine and compare the lookup table sizes of various lookup-table based refractive forward projection methods.

calculate the lookup table size for a common specific scenario.

For the complexity classes, we compared how much the lookup table size grows with the maximum selected object distance $\hat{z}_p$, the radial back projection accuracy a and the distance z_p^* up to which the accuracy a is achieved (see Fig. 5.3). The accuracy a is defined as $a = 1/\hat{e}_{p,r}$, where $\hat{e}_{p,r}$ denotes the maximum radial back projection error (see Fig. 5.3).

For the specific test scenario, we choose a setup, which is similar to the requirements in a real-world ship hull inspection scenario. We determine the lookup table size, which is required to achieve a maximum back-projection error of $\hat{e}_{p,r} = 0.01\,\text{mm}$ in a viewing volume up to a distance of $\hat{z}_p = 1\,\text{m}$ and a distance $z_p^* = 1\,\text{m}$ with 2 cameras and three wavelengths.

5.4.1. Maas' 2D lookup table

The input parameters of the 2D lookup table of Maas [1] are the indices of the discretised coordinates of the plane of incidence. The size n_{lut} of the lookup table is equivalent to the number of object points in the 2D point grid being mapped. n_{lut} depends on the viewing area $A \subset \mathbb{R}^2$ in the plane of incidence and on the accuracy a required. The viewing area A is limited by the minimum object distance $d + t$ in front of the port, by

the maximum object distance $\hat{z}_p$, and by the field of view of the camera in water.

Similar to a perspective camera, the maximum radial distance of the viewing area A grows linearly with the maximum object distance $\hat{z}_p$. Assuming a uniform discretisation of A in the r and z directions, the number n_{lut} of discrete grid points or lookup table entries, respectively, grows quadratically

$$n_{lut}(\hat{z}_p) \in O(\hat{z}_p^2). \tag{5.14}$$

Similarly, a linear increase in accuracy results in a quadratic increase in grid points. The lookup table size grows quadratically with a, that is,

$$n_{lut}(a) \in O(a^2). \tag{5.15}$$

Due to the uniform discretisation of the plane of incidence, the accuracy of Maas's method does not vary significantly with the object distance and, thus, has the complexity

$$n_{lut}(z_p^*) \in O(1). \tag{5.16}$$

As shown in (5.14) and (5.15), the main disadvantage of Maas's method lies in the memory requirement of the lookup table for large visual ranges and high accuracies. Like for all lookup table-based methods described in this chapter, it also multiplies with the number of wavelengths and cameras used.

In order to compare the lookup table size n_{lut} of Maas's method with alternative methods having the same accuracy $a = 1/\hat{e}_{p,r}$, the point grid resolution $(\Delta r, \Delta z)$ need to be determined. The optimal solution for $(\Delta r, \Delta z)$ with minimal lookup table size n_{lut} is given by

$$\begin{aligned}(\Delta r, \Delta z)^* &= \underset{(\Delta r, \Delta z)\in\mathbb{R}^2}{\arg\min}\ f_c(\Delta r, \Delta z),\\ f_c(\Delta r, \Delta z) &= \left| \max_{^{poi}\boldsymbol{p}\in A} \left(\left| e_{p,r}(^{poi}\boldsymbol{p}, \Delta r, \Delta z) \right| \right) - \hat{e}_{p,r} \right|,\end{aligned} \tag{5.17}$$

where $^{poi}\boldsymbol{p} = (r_p, z_p)^T$ denotes a point in the viewing area A and where $e_{p,r}()$ describes the radial refractive back projection error

$$e_{p,r}(^{poi}\boldsymbol{p}, \Delta r, \Delta z) = r'_p - r_p. \tag{5.18}$$

The radial coordinate r'_p of the ray in water at the object distance z_p is calculated by

$$\begin{aligned} r'_p &= r'_{pw} + (z_p - d + t)\frac{r_{d,w}}{\sqrt{1 - r_{d,w}^2}}, \\ (r'_{pw}, r_{d,w}) &= f_{rb}(r'_{ap}). \end{aligned} \tag{5.19}$$

Here, $f_{rb}()$ denotes the refractive back projection (5.8) and r'_{ap} stands for the result of Maas' refractive forward projection method

$$\begin{aligned} f_{rf,maas} : \mathbb{R}^2 &\to \mathbb{R}, \\ {}^{poi}\boldsymbol{p} &\mapsto r'_{ap}, \end{aligned} \tag{5.20}$$

using the grid resolution $(\Delta r, \Delta z)$. Note that the result of $f_{rf,maas}$ is based on the interpolation of four grid points to increase the accuracy and to reduce the size of the lookup table.

5.4.1.1. Results

For our wide-angle underwater camera system (see Table 2.1) with a back projection accuracy $a = 1/0.01\,\text{mm}$ up to the distance $z_p^* = 1\,\text{m}$ and refractive indices as described in Table 6.2, the optimisation resulted in a grid of points spaced by about 1.7 mm in r-direction and by 1.5 mm in z-direction. This corresponds to a minimum lookup table size $n_{lut} \approx 160,000$ or a memory requirement of ∼626 kB using single precision floating point numbers. The result shows, that in spite of the twofold quadratic dependency with respect to the viewing distance (5.14) and back projection accuracy (5.15), the memory requirement of Maas's 2D lookup table-based method is relatively low and well appropriate for accurate, close range applications. But for multiple wavelengths, multiple cameras, higher accuracies, and larger object distances huge amounts of lookup table memory are needed.

In Fig. 5.4, the memory size required for three different wavelength (e.g. red, green, blue) and two cameras (stereo) at different maximum object distances $\hat{z}_p$ is shown. The memory required by Maas's 2D lookup tables grows rapidly with $\hat{z}_p$. At a distance of 5 m, the lookup table requires ∼89 MB , at 10 m ∼355 MB, at 20 m ∼1.4 GB and at 50 m already ∼8.8 GB. These values would increase even further with higher accuracies.

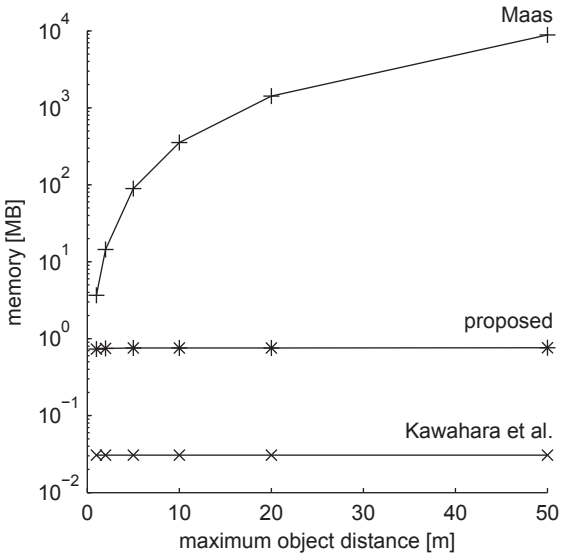

Figure 5.4.: Memory requirements of different lookup table-based methods as a function of the maximum object distance $\hat{z}_p$ using 2 cameras (stereo) and 3 wavelengths (red, green, blue) for a radial back projection accuracy of 0.01 mm

5.4.2. Kawahara et al.'s 1D lookup table

The recently presented iterative method of Kawahara et al. [2] calculates refractive forward projection by a recurrence relation and a one-dimensional lookup table. Using the recurrence relation, the radial coordinate $\tilde{r}_{pw,i}$ of the refrax on the port-water interface is iteratively adjusted until it converges against the correct solution r'_{pw}. The recurrence relation is based on the idea of the repetitive alteration of the position $\tilde{z}_f$ of the virtual focal point with a resulting repetitively changing perspective projection of the object point on the port-water interface. The lookup table contains the precomputed unique $r'_{pw,j} \mapsto z_{f,j}$ mapping, which describes the correspondence between the correct radial coordinates $r'_{pw,j}$ of the refraxes on the port-water interface and the z-coordinates $z_{f,j}$ of the corresponding virtual focal points. At each iteration step i, the object point is projected by perspective projection on the port-water interface at the radial coordinate $\tilde{r}_{pw,i}$ using the current estimate $\tilde{z}_{f,i}$ of the virtual focal point. In the next step, the perspective projection ($\tilde{r}_{pw,i+1}$) is performed again, however, this time with the new, modified focal length $\tilde{z}_{f,i+1}$ obtained from the lookup table using the last result of the perspective projection $\tilde{r}_{pw,i}$. The recurrence relation $\tilde{r}_{pw,i} \mapsto \tilde{r}_{pw,i+1}$ corresponds to a monotonic sequence, which converges finally against the correct solution. Therefore, the iteration process can stop as soon as the change in $\tilde{r}_{pw,i}$ is sufficiently small.

The lookup table in Kawahara et al.'s recurrence method requires a certain size n_{lut}

to achieve the minimum radial refractive back projection accuracy a up to the object distance z_p^*. Due to the perspective projection of object points, the size of Kawahara et al.'s method is independent of the maximum visual distance

$$n_{lut}(\hat{z}_p) \in O(1). \tag{5.21}$$

Furthermore, the resolution or the size of the lookup table, respectively, grows only linearly with the radial back projection accuracy $a = 1/\hat{e}_{p,r}$

$$n_{lut}(a) \in O\left(a\right), \tag{5.22}$$

and linearly with the visual distance z_p^* up to which the accuracy is achieved

$$n_{lut}(z_p^*) \in O(z_p^*). \tag{5.23}$$

The actual required lookup table size n_{lut} of Kawahara et al.'s method is determined by finding the optimal resolution $\Delta r'_{pw}$ of the discrete radial coordinates on the port-water interface. Similar to (5.17), we have

$$\begin{aligned} \Delta r'^{*}_{pw} &= \underset{\Delta r'_{pw} \in \mathbb{R}}{\arg\min}\, f_c(\Delta r'_{pw}), \\ f_c(\Delta r'_{pw}) &= \left| \max_{^{poi}\boldsymbol{p} \in A} \left(\left| e_{p,r}(^{poi}\boldsymbol{p}, \Delta r'_{pw}) \right| \right) - \hat{e}_{p,r} \right|, \end{aligned} \tag{5.24}$$

where $e_{p,r}(^{poi}\boldsymbol{p}, \Delta r'_{pw})$ corresponds this time to the forward projection by Kawahara et al.'s method using the resolution $\Delta r'_{pw}$.

5.4.2.1. Results

For our setup with one camera and one wavelength (see section 5.4.1.1), $\Delta r'_{pw} \approx 0.028\,\text{mm}$ resulting in a lookup table size of only 1336 elements corresponding to a memory consumption of $\sim$5.2 kB using 4 byte floating point numbers. Herewith, Kawahara et al.'s method requires the least amount of memory of all lookup table-based methods evaluated. The memory requirement is significant two orders of magnitudes lower than for the other methods, but as shown in section 5.7, the algorithm is also the slowest lookup table-based refractive forward projection method.

5.5. Proposed refractive forward projection methods

5.5.1. Proposed back projection-based refractive forward projection method

If the underwater camera needs to be calibrated or the indices of refraction change quickly on-site, lookup table-based refractive forward projection methods are not eligible. For such cases, we propose a more efficient method, which is based on the classic back projection-based and Newton's method-based refractive forward projection algorithm.

Our proposed improvements consist, firstly, of a reduction of the number of calculations of reciprocal square roots, secondly, of replacing the standard reciprocal square root method by a faster algorithm, and, thirdly, of a better initial guess. Our modifications lead to an overall reduction in execution time of 42 % compared to Kawahara et al.'s [2] implementation of the algorithm used for benchmarking.

Classic back projection-based refractive forward projection methods work as follows. Starting with an initial direction of the light ray in air, the direction is iteratively changed until the corresponding refracted light ray in water is close enough to the according object point to be projected. For computational efficiency and numerical stability, the calculation is carried out in the plane of incidence and uses only radial coordinates. Kawahara et al. [2], for example, computed the radial refractive back projection error (see Fig. 5.3) and minimised it using Newton's method [86].

Newton's method requires the definition of the error function and the definition of its derivative. In Kawahara et al.'s implementation the radial refractive back projection error is used (Fig. 5.3, (5.19)). The error function $e_{p,r}(r'_u, {}^{poi}\boldsymbol{p})$ and its derivative $\partial e_{p,r}(r'_u, {}^{poi}\boldsymbol{p})/\partial r'_u$ consist of additions, multiplications and inverses of reciprocal square roots $1/\sqrt{x}$. Divisions were avoided due to the choice of the parameter r'_u, which describes the radial coordinate of the light ray in air at the distance $z = 1$.

The term $1/\sqrt{x}$ represents by far the computationally most expensive operation in the computation of the error function and of its derivative [87]. Kawahara et al.'s method performs six $1/\sqrt{x}$ calculations per iteration step. However, only two are necessary. Our

Listing 5.1: A fast variant of $1/\sqrt{x}$: A 64 bit version with the optimized *magic number* of Robertson [89]

```
double rsqrt64(double x){

  // Initial guess with magic constant
  uint64_t i= *(uint64_t*)&x;
  i=0x5fe6eb50c7b537a9-(i>>1);
  double y=*(double*)&i;

  // Increase of accuracy by
  // Newton's method, 2 iterations
  double x2=x*0.5;
  y=y*(1.5-(x2*y*y));
  y=y*(1.5-(x2*y*y));

  return y;
}
```

simplified, computationally more efficient equations read as follows:

$$
\begin{aligned}
e_r(r'_u, {}^{poi}\boldsymbol{p}) &= r'_u(d + n_a(tb_p + bb_w)) - r_p, \\
b &= z_p - d - t, \\
b_i &= \frac{1}{\sqrt{n_i^2 + (n_i^2 - n_a^2)r'^2_u}},\ i \in \{p, w\}, \qquad (5.25) \\
\frac{\partial e_r(r'_u, (r_p, z_p)^T)}{\partial r'_u} &= d + n_a\left(tn_p^2 b_p^3 + bn_w^2 b_w^3\right), \qquad (5.26)
\end{aligned}
$$

where ${}^{poi}\boldsymbol{p} = (r_p, z_p)^T$.

The computation of the reciprocal square root still contributes significantly to the refractive forward projection's overall execution time and is worth being optimised. Therefore, we replace the standard implementation by a faster variant. The more efficient algorithm used to compute the reciprocal square root was firstly published in 2002. Until today, the origins and the author of the fast reciprocal square root are unknown. A first implementation of the original version of the fast $1/\sqrt{x}$ appeared in the source code of the computer game Quake III Arena [88], whose binaries were released in 1999. In numerous publications the way of functioning was analysed and new improvements were presented.

The principle structure of the fast $1/\sqrt{x}$ is shown in Listing 5.1. The algorithm consists

of two parts, in which, at first, the initial guess is computed, and then Newton's method is applied to increase the algorithm's accuracy. The significant improvement in speed of the algorithm is achieved by the efficient computation of the initial guess. In comparison to computationally more expensive floating point arithmetic, computationally cheaper integer operations in combination with the magic number are directly applied on the floating point input argument. For a more detailed description of this step, we refer to [89].

We have used the 64-bit magic number of Robertson [89], which minimises the maximum relative error after one Newton iteration. However, due to our required radial refractive back projection accuracy of 0.01 mm, the extension with a second iteration step was necessary.

Despite the loss in speed due to the second Newton iteration step, our measurements showed that the algorithm is still faster than the alternative (Streaming SIMD Extensions (SSE)) CPU hardware-based approach. We also took advantage of the parallel computational capability of the processor (single instruction, multiple data (SIMD)) to calculate two reciprocal square roots in parallel, but the proposed method was still faster. The accuracy of the in hardware implemented reciprocal square root is also relatively inaccurate so that here also two Newton iterations are necessary.

Furthermore, we shortened the execution time of Kawahara et al.'s [2] back projection-based refractive forward projection method by using a more accurate initial guess. Kawahara et al. used the perspective projection $\tilde{r}'_{u,0} = r_p/z_p$ of the object point $(r_p, z_p)^T$. But due to the inherent magnification caused by refraction, the radial distance of the actual refractive projection is always larger. This was described in chapter 4 in more detail and was visualised in Fig. 4.6. It is advantageous to start with a larger (magnified) initial guess. The selection of $\tilde{r}'_{u,0}$ is system specific and not completely independent of the configuration of the underwater camera. Our empirical studies showed that for our system, a constant magnification of 1.45 seems to be a good value to reduce the number of iteration steps and, thus, the algorithm's execution time. The new initial guess reads as follows

$$\tilde{r}_{u,0} = 1.45\frac{r_p}{z_p}. \tag{5.27}$$

Depending on the application, further optimisations of the initial guess are possible. If only the indices of refraction vary and the geometry of the camera is known, using a constant lookup table for the initial guess can makes sense. If the alteration speed

of the refractive indices is low and the number of refractive forward projections is high, a dynamically updated lookup table-based refractive forward projection method may represent a more efficient alternative. As the discussed alternative methods are more application specific, we focus in the rest of the chapter only on the variant (5.27) with a constant magnification.

5.5.2. Proposed 2D lookup table-based refractive forward projection method

If the underwater camera's intrinsics and the indices of refraction are known and constant, the forward projection can be discretised, precalculated, and stored in a lookup table. The 2D lookup table-based method of Maas [1] belongs to this category and represents in this chapter the fastest refractive forward projection algorithm. As described in section 5.4.1, however, the size of Maas's lookup table grows quadratically with the object distance. We propose an alternative 2D lookup table-based approach, which is slightly slower, but uses a lookup table whose size does not grow at large visual ranges, requires significantly less memory and is even considerably more accurate at close visual ranges. Our method achieves significant savings in memory by taking advantage of refractive effects.

As discussed in chapter 4, the refractive magnification function $f_m(r', z_p)$ describes the ratio r'/r of the radial distances of the refractive forward projection r' and the corresponding perspective projection r of an object point ${}^{poi}\boldsymbol{p} = (r_p, r_z)^T$ on a common r-axis parallel line in air in the plane of incidence. As was explained in chapter 4 and illustrated in Fig. 4.6, $f_m(r', z_p)$ shows a strong object distance dependency for small object distances z_p and converges for large z_p against a function, which is independent of z_p. We exploit this property for the implementation of our 2D lookup table.

Our proposed lookup table corresponds to the mapping

$$\begin{aligned} f_{lut} : \mathbb{N}^2 &\to \mathbb{R}, \\ (i, j) &\mapsto r'_{i,j} \end{aligned} \tag{5.28}$$

in which the indices i of the discrete radial distances r_i obtained by perspective projection and the indices j of the discrete object point distances $z_{p,j}$ are mapped to the corresponding radial coordinates r'_i of refractive projections.

In contrast to Maas's approach, the distribution of the $z_{p,j}$ being mapped is not uniform. Instead, we adapt the $z_{p,j}$ distribution according to the distance dependency of

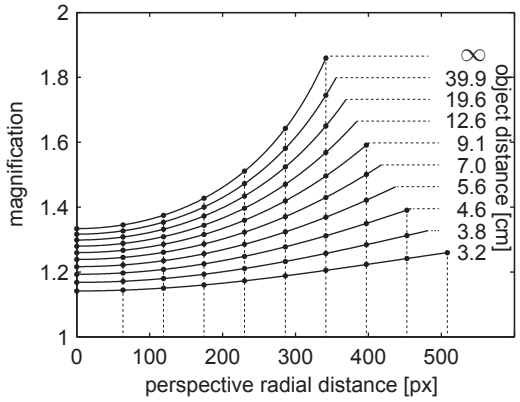

Figure 5.5.: Exemplary discretisation of the plane of incidence used in our memory and time efficient 2D lookup table approach: By discretising the plane of incidence in r-direction with a decreasing resolution proportional to z, and in z-direction by a non-uniform object point distribution, the size of the 2D lookup table becomes independent of the maximum object distance. For the discretisation of the z-direction, the uniform discretisation of the codomain of the magnification function $f_m(r, z)$ is used.

$f_m(r', z_p)$. That is, the density of the $z_{p,j}$ for small object distances is highest, decreases with larger object distances and requires only one entry for very large object distances. The same property applies to $f_m(r, z_p)$, where f_m now depends on the radial coordinate of the perspective projection r instead of r'. An illustration of the distance dependency of $f_m(r, z_p)$ is given in Fig. 5.5.

To obtain a useful non-uniform discretisation of the object distances $z_{p,j}$, we use the results obtained in chapter 4. For normal port thickness, the difference in refractive magnification between two object points with the same radial coordinate but different distances is greatest at large radial distances (see Fig. 4.6a). Therefore, we consider $f_m(\hat{r}', z_p)$ at the maximum radial distance $\hat{r}'$. The non-uniform distribution of the object distances is achieved by uniformly discretising the value range between the minimum magnification $\check{f}_m = f_m(\hat{r}', \check{z}_p)$ and maximum magnification $\hat{f}_m = f_m(\hat{r}', \hat{z}_p)$ whereby the resulting $f_{m,j}$ are equally spaced by Δf_m.

For the initialisation of the lookup table, the relation between the discrete magnification $f_{m,j}$ and the corresponding discrete object distances $z_{p,j}$ needs to be known. By

rearranging the equation (4.7) of the magnification function, we have

$$\begin{aligned} z_{p,j} &= \frac{f_{m,j} b}{1 - f_{m,j} b_w}, \\ f_{m,j} &= \Delta f_m j + \check{f}_m, \\ b &= (b_p - b_w)t + (1 - b_w)d. \end{aligned} \tag{5.29}$$

In our method, the second argument of the lookup table corresponds to the index of the discrete radial coordinate r_i of a perspective projection. The distribution of the radial coordinates differs from Maas's straight forward lookup table method and is similar to Kawahara et al.'s [2] approach. For the initialisation of the lookup table, the mapping $(i, j) \mapsto r_{p,ij}$ is required, which is

$$r_{p,ij} = \frac{z_{p,j} \Delta r}{f} i, \tag{5.30}$$

where Δr denotes the distances between the discrete r_i.

Determining the actual radial coordinate $r' \in \mathbb{R}$ of the refractive forward projection of an object point $(r_p, z_p)^T$ using our lookup table requires two steps. Firstly, the object point coordinates $(r_p, z_p)^T$ are mapped to the real lookup table indices $(i', j') \in \mathbb{R}^2$

$$\begin{aligned} (r_p, z_p)^T &\mapsto (i', j'), \\ i' &= \frac{|r_p| f}{z_p \Delta r}, \\ j' &= \frac{f_m(z_p) - \check{f}_m}{\Delta f_m}, \end{aligned} \tag{5.31}$$

Using the constants b and b_w from the lookup table's initialisation (5.29), the equation of $f_m(z_p)$ simplifies to

$$f_m(z_p) = \frac{z_p}{b + b_w z_p}. \tag{5.32}$$

The second step corresponds to a standard linear interpolation having the form

$$r' = \operatorname{sign}(r_p)(\boldsymbol{M} \boldsymbol{v}_1)^T \boldsymbol{v}_2 \tag{5.33}$$

in which the lookup table values of the four discrete neighbours of (i', j') are determined

$$\boldsymbol{M} = \begin{pmatrix} f_{lut}(\lfloor i' \rfloor, \lfloor j' \rfloor) & f_{lut}(\lceil i' \rceil, \lfloor j' \rfloor) \\ f_{lut}(\lfloor i' \rfloor, \lceil j' \rceil) & f_{lut}(\lceil i' \rceil, \lceil j' \rceil) \end{pmatrix} \tag{5.34}$$

and subsequently weighted by

$$\begin{aligned} \boldsymbol{v}_1 &= (a_1, 1 - a_1)^T, \\ \boldsymbol{v}_2 &= (a_2, 1 - a_2)^T \end{aligned} \tag{5.35}$$

according to their distances to (i', j')

$$\begin{aligned} a_1 &= \lceil i' \rceil - i', \ a_1 \in \mathbb{R}, \\ a_2 &= \lceil j' \rceil - j', \ a_2 \in \mathbb{R}. \end{aligned} \tag{5.36}$$

Here, $\lfloor \rfloor$ and $\lceil \rceil$ denote the floor and ceiling functions to round a real number to the next smallest or largest integer number, respectively.

The radial refractive back projection accuracy of our 2D lookup table method is highest at close proximity to the camera and decreases with growing object distance. This is a result of the perspective discretisation of the radial coordinates. The density of discrete object points in the plane of incidence in radial direction decreases with increasing object distance. The accuracy distribution in the plane of incidence of our approach is, thus, similar to the method of Kawahara et al. [2] and different to the method of Maas [1] in which the accuracy is approximately uniformly distributed in the plane of incidence. Consequently, the accuracy of our and Kawahara et al.'s method is best at close visual ranges, whereas Maas's method is more accurate at large object distances but also requires significantly more memory.

In practice, the lower accuracy at large object distance should not be disadvantageous. For stereo-based 3D reconstruction applications, for example, the reconstruction accuracy of stereo cameras decreases with object distance, too, so that the actual effect on the 3D reconstruction error should be negligible.

In terms of the memory complexity of the lookup table, our approach behaves similarly to the method of Kawahara et al.. Due to the discretisation of the radial coordinates of the perspective projections of object points and the non-uniform discretisation of the

object distances, the method is independent of the maximum object distance $\hat{z}_p$

$$n_{lut}(\hat{z}_p) \in O(1). \tag{5.37}$$

The lookup table size grows only linearly with the radial back projection accuracy $a = 1/\hat{e}_{p,r}$

$$n_{lut}(a) \in O(a), \tag{5.38}$$

and linearly with the visual distance z_p^* up to which a is guaranteed

$$n_{lut}(z_p^*) \in O(z_p^*). \tag{5.39}$$

The optimal distance Δr between discretised perspective radial coordinates and the optimal number n_z of discretisations in z-direction is formulated by

$$\begin{aligned}(\Delta r, n_z)^* &= \underset{(\Delta r, n_z)\in\mathbb{R}\times\mathbb{N}}{\arg\min}\ e_{p,r}(\Delta r, n_z),\\ e_{p,r}(\Delta r, n_z) &= \left| \max_{{}^{poi}\boldsymbol{p}\in A} \left(\left| e_{p,r}({}^{poi}\boldsymbol{p}, \Delta r, n_z) \right| \right) - \hat{e}_{p,r} \right|,\end{aligned} \tag{5.40}$$

5.5.2.1. Results

For the setup given (see section 5.4.1.1), our optimisation method results in $(\Delta r, n_z) \approx (1.1\,\text{px}, 85)$. The lookup table has ∼32074 entries and requires ∼125 kB using 4 byte floating point numbers. The result demonstrates the significant reduction in the size of the lookup table. In comparison to Maas's method, our proposed 2D lookup consumes ∼80 % less memory.

5.6. Proposed evaluation of polynomial equation-based refractive forward projection methods

Before carrying on with the main benchmark, we will propose evaluations of another two forward projection methods to address the forward projection problem by finding real roots of polynomial equations. Glaeser and Schröcker [83] and Agrawal et al. [73] showed that the forward projection with one interface can be described by a 4th degree polynomial and the forward projection with two parallel interfaces by a polynomial

of degree 12. Whereas closed-form solutions for polynomials of degrees up to 4 are known, solving polynomials with a higher degree requires iteration based-methods, in general. For benchmarking the single interface forward projection problem with a single interface [83], we used the implementation [90] of Herbison-Evans's [85] closed-form solution. For the refractive forward projection with two parallel interfaces [73], we compared the polynomial root finder of Francis [91] based on the balanced-QR reduction of the companion matrix, and the root finder of Jenkins and Traub [92]. The according implementations of the algorithms can be found in [93] and [94].

5.6.1. Results

Solving Agrawal et al.'s 12th degree polynomial using standard polynomial root finders is only possible up to a certain object distance. In the correct case, the polynomial root solvers find all existing polynomial roots from which, subsequently, the correct real root with a very small radial back projection error is selected. Our experiments showed however, that all polynomial root solvers tested suffer from missing roots at distances of more than $\sim$10 m. We hypothesis that this is caused by the limited accuracy of the root solvers' underlying numerical system.

5.7. Proposed benchmarking of various refractive forward projection methods

In our experiment, we benchmarked the whole refractive forward projection chain (5.6) including the transformation from object space coordinates to plane of incidence coordinates, the transformation vice versa, and the perspective camera projection. To measure the execution times of the refractive forward projection methods with different object distances and angles, the 3D viewing volume of the camera was discretised. We used four different layers of point grids at the distances 0.25, 0.5, 0.75 and 1 m with a uniform resolution of 1 point per mm in x- and y-direction. For a fair comparison with the same maximum radial refractive back projection error $\hat{e}_{p,r} = 0.01$ mm within the viewing range of $d + t = 3.2$ cm to $z_p^* = 1$ m, the maximum radial back projection error of the back projection-based approaches and the lookup table size of the lookup table-based approaches were set accordingly (see 5.4.1, 5.4.2, 5.5.2). The accuracy of the polynomial root solvers could not be adjusted with respect to the accuracy criterion. For accurate time measurement, every discrete object point was projected a hundred times and the

execution times of the ∼230 million projections per method averaged. All algorithms were implemented in C/C++ and executed on an Intel Core i7-2600 CPU.

5.7.1. Results

The refractive forward projection methods benchmarked can be grouped into lookup table-based and non-lookup table-based methods. Lookup table-based methods are applicable to refractive forward projection problems with arbitrary interface numbers, but require a calibrated camera and known, constant refractive indices. In contrast to that, non-lookup tabled-based methods are more flexible and allow dynamic changes in the optical system. The measured execution times are listed in Table 5.1.

5.8. Conclusion

As our experimental measurements show, the two most efficient refractive forward projection methods are members of the lookup table category. Maas's [1] straight forward 2D lookup table approach is fastest (105 ns) due to its limited complexity, followed by our proposed 2D lookup table method (+11 %), whose lookup table size is independent of the maximum viewing range and which is more accurate at close distances. The iterative 1D lookup table-based forward projection method of Kawahara et al.'s [2] is relatively slow (+83 %). It is outperformed by our proposed back projection based methods for one (+59 %) or two interfaces (+63 %), which use a more efficient computation of the back projection error and a better initial guess. Our back projection variant for two interfaces was described in section 5.5.1. The single interface variant corresponds to a specialisation of the two interface case and uses less terms in the computation of the error function (5.25) and of its derivative (5.26). Compared with the implementation of the classic back projection-based method by Kawahara et al. [2] (+190 % with respect to Maas's [1] method), the benchmark demonstrates the significant savings in time achieved by our modifications. The polynomial-based forward projection methods lag far behind. Whereas the execution time of Glaeser and Schröcker's [83] closed-form solution still has the same order of magnitude (+441 %), the computation of Agrawal et al.'s [73] 12th degree polynomial approach requires significantly more time and is two orders of magnitude slower than the reference method of Maas [1] using the polynomial root solver of Jenkins and Traub [92] (+11,710 %) or the polynomial root solver of Francis [91] (+30,471 %).

Table 5.1.: Benchmark: Refractive forward projection ($\mathbb{R}^3 \rightarrow \mathbb{R}^2$)

Solution method	Application	Time [ns]
1 interface		
1. Closed-form solution [85] of a quartic equation [83]	(V)	568
2. NM, BP; 1× fast $1/\sqrt{}$; guess: perspective, const. magnification (proposed)	V	167
2 interfaces		
3. 12th degree polynomial [73] with a polynomial root solver [92]/[91]	(V)	12,400/32,100
4. NM, BP, 6× $1/\sqrt{}$, guess: perspective [2]	(V)	304
5. NM, BP; 2× $1/\sqrt{}$, fast $1/\sqrt{}$; guess: perspective, const. magnification (proposed)	V	171
n interfaces		
6. Reccurence, 1D LUT, guess: perspective, LUT size: 5.2 kB [2]	(C)	192
7. 2D LUT, distance independence, linear interpolation, LUT size: 125 kB (proposed)	C	117
8. 2D LUT, linear interpolation, LUT size: 626 kB [1]	C; limited visibility	105

V: uncalibrated camera or varying refractive indices; C: calibrated camera and constant refractive indices
(): Slower than fastest method for this application
NM: Newton method; BP: back projection; LUT: lookup table

6. Minimising projection and reconstruction errors using accurate system knowledge

6.1. Introduction

Errors in image projection and reconstruction due to inaccuracies, dynamic changes or even inadequacies in the refractive model parameters of (stereo) cameras with a flat port have not been explicitly investigated yet. But determining the relevant model parameters of the refractive system and their effects is essential for very accurate imaging and photogrammetry. For that reason, we investigate in section 6.4 the errors arising due to inaccurate modelling of the port, and in section 6.5 the errors caused by inaccurate indices of refractions. In order to evaluate the errors in 2D imaging and 3D reconstruction, we start this chapter with two preliminary sections in which we propose at first a method for accurate 3D reconstruction given a single point-pair (section 6.2), and secondly the computation of the dominating wavelength from the camera sensor's data sheet (section 6.3).

6.2. Refractive 3D reconstruction from a single point correspondence

6.2.1. Introduction

Underwater 3D reconstruction with a flat port underwater stereo camera differs from standard in-air stereo reconstruction. For classic rectified images captured in-air, the epipolar line constraint is applicable and object points can be reconstructed by their 2D coordinates in one image and their disparity value in the second image. For underwater images however, a complete undistortion is not straightforward and requires the depth

of each pixel. Therefore, 3D reconstruction is based on distorted underwater images, where 4 instead of 3 parameters are given, namely the two 2D image points, $\boldsymbol{u}_l$ and $\boldsymbol{u}_r$, in the left and right camera image

$$\begin{aligned} f_{rr} :\ & \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^3, \\ & (\boldsymbol{u}_l, \boldsymbol{u}_r) \mapsto (\boldsymbol{p}). \end{aligned} \tag{6.1}$$

In this overdetermined system, the two rays in water, $(\boldsymbol{p}_{pw,l}, \boldsymbol{d}_{w,l})$ and $(\boldsymbol{p}_{pw,r}, \boldsymbol{d}_{w,r})$ of the corresponding back projections f_{rb} of the image points do not necessarily intersect in one point if the modelled refractive system is inaccurate or subject to image noise (see Fig. 6.1).

Only few publications, which deploy the physics-based refractive camera model, elaborate on stereo 3D reconstruction from a singe point correspondence.

In 1997, based on least squares, Li et al. [95] determined the object point to be reconstructed by minimising the distance of the object point to the back-projected rays of corresponding points in the left and right camera image. A similar approach for refractive multi view geometry was later also described by Jordt [78]. Li et al.'s method is not as accurate as our proposed method, as it does not minimise the error in the image space where the points where measured.

6.2.2. Proposed 3D reconstruction

Our algorithm, minimises the reprojection error in image space, and thus is more accurate, as the measurement noise is accurately modelled by normal distributed Gaussian noise. Our proposed initial guess computation is similar to the object point calculated by Li et al.. However, we formulate the initial guess as a set of linear equations.

In our proposed method, we model 3D reconstruction by minimising the reprojection error. Using the refractive forward projection function f_{rf}, the reconstructed object point $\boldsymbol{p}$ is defined by

$$\boldsymbol{p} = \underset{\boldsymbol{p} \in \mathbb{R}^3}{\arg\min} \sum_{i \in \{l,r\}} \left| f_{rf}(\boldsymbol{p}, i) - \boldsymbol{u}_i \right|^2 . \tag{6.2}$$

The equation is solved by iterative numerical minimisation. A good initial guess of the reconstructed point, represents the point $\boldsymbol{p}_0$, which is closest to the two rays in water (see Fig. 6.1). $\boldsymbol{p}_0$ is located between the two ray points $\boldsymbol{p}_l$ and $\boldsymbol{p}_r$,

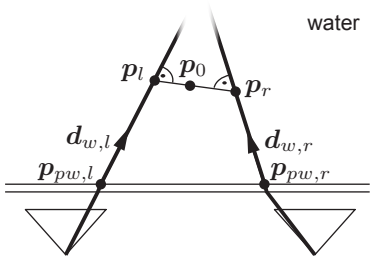

Figure 6.1.: 3D reconstruction with a flat port underwater stereo camera: In general, the rays $(\boldsymbol{p}_{pw,l}, \boldsymbol{d}_{w,l})$ and $(\boldsymbol{p}_{pw,r}, \boldsymbol{d}_{w,r})$ of the corresponding back-projected image points of the left and right camera image do not intersect and a non-linear numerical optimisation is required. A good initial guess is the point $\boldsymbol{p}_0$, which is closest to the two rays.

$$\boldsymbol{p}_0 = \frac{\boldsymbol{p}_l + \boldsymbol{p}_r}{2}, \tag{6.3}$$

which lie per definition on the ray

$$\begin{aligned} \boldsymbol{p}_l &= \boldsymbol{p}_{pw,l} + s_l \boldsymbol{d}_{w,l},\ s_l \in \mathbb{R}, \\ \boldsymbol{p}_r &= \boldsymbol{p}_{pw,r} + s_r \boldsymbol{d}_{w,r},\ s_r \in \mathbb{R}, \end{aligned} \tag{6.4}$$

and are the ray points with the smallest distance to each other. Therefore, the line connecting $\boldsymbol{p}_l$ and $\boldsymbol{p}_r$ is perpendicular to both rays

$$\begin{aligned} \boldsymbol{d}_{w,l}^T (\boldsymbol{p}_r - \boldsymbol{p}_l) &= 0, \\ \boldsymbol{d}_{w,r}^T (\boldsymbol{p}_r - \boldsymbol{p}_l) &= 0. \end{aligned} \tag{6.5}$$

The initial guess $\boldsymbol{p}_0$ is obtained by solving this linear equation system using (6.3), (6.4) and (6.5). Here, the unknown variables s_l and s_r can be calculated by substituting (6.4) into (6.5), which results in the following linear equation:

$$\underbrace{\begin{pmatrix} -\boldsymbol{d}_{w,l}^T \boldsymbol{d}_{w,l} & \boldsymbol{d}_{w,l}^T \boldsymbol{d}_{w,r} \\ -\boldsymbol{d}_{w,l}^T \boldsymbol{d}_{w,r} & \boldsymbol{d}_{w,r}^T \boldsymbol{d}_{w,r} \end{pmatrix}}_{\boldsymbol{A}} \begin{pmatrix} s_l \\ s_r \end{pmatrix} = \underbrace{\begin{pmatrix} \boldsymbol{d}_{w,l}^T \\ \boldsymbol{d}_{w,r}^T \end{pmatrix} (\boldsymbol{p}_{pw,l} - \boldsymbol{p}_{pw,r})}_{\boldsymbol{b}} \tag{6.6}$$

And thus

$$\begin{pmatrix} s_l \\ s_r \end{pmatrix} = \boldsymbol{A}^{-1} \boldsymbol{b}. \tag{6.7}$$

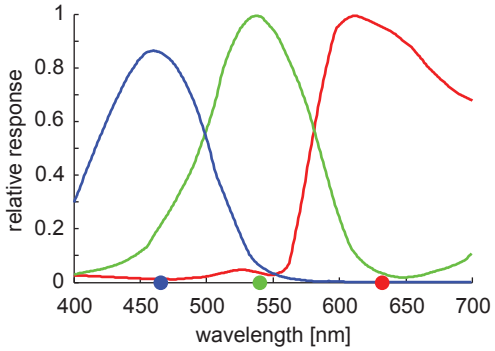

Figure 6.2.: Spectral sensitivity characteristic curves $S_i(\lambda)$ of the camera sensor used. The coloured circles show the dominating wavelength λ_r, λ_g and λ_b of each colour channel assuming a uniform wavelength distribution (white light).

6.3. Computation of the dominating wavelengths

6.3.1. Introduction

Before light is sensed by the imager of an underwater camera, a complex light transmission process takes place. The light spectrum sensed depends on the spectrum of the light source, the radiance of the objects imaged, and absorption, forward and back scattering in water [96]. Once light finally reaches the camera housing, it is then further filtered by the port, the lens, the infrared cut-off filter behind the lens and the colour filter array in front of the imager. In general, the colour filter is a Bayer filter array consisting of 3 different colour filters for red, green and blue light. The spectral sensitivity characteristics of the colour image sensor of our camera [53] are depicted in Fig. 6.2.

Due to the band-pass-like spectral sensitivity characteristics of the colour channels of the image sensor, accurate measuring of all occurring wavelengths of the light spectrum is not possible. Due to complex light transmission process from the light source to the imager, determining only the dominating wavelengths λ_r, λ_g and λ_b of each colour channel is not straight forward, either.

6.3.2. Proposed wavelength computation

For simplification, we approximated the dominating wavelength by neglecting the effects before the light passes the colour filters and directly derived λ_r, λ_g and λ_b from the

spectral characteristics of the image sensor using

$$\lambda_i = \frac{\int \lambda S_i(\lambda) d\lambda}{\int S_i(\lambda) d\lambda}, \quad i \in \{r, g, b\}, \tag{6.8}$$

where $S_i(\lambda)$ denotes the spectral sensitivity.

Applying (6.8) on the spectral characteristics of our camera sensor, we obtain for the three dominating wavelengths $\lambda_r = 630\,\text{nm}$, $\lambda_g = 538\,\text{nm}$ and $\lambda_b = 463\,\text{nm}$.

6.4. Effects of inaccurate port modelling on projection and reconstruction

6.4.1. Introduction

Prior literature differs in how accurate the pose and the thickness of a port of an underwater camera need to be known. In the publications of Treibitz et al. [68, 71], for instance, neither the tilt nor the thickness of the port were modelled. Compared to this, other authors explicitly calibrated the relative position and orientation between the interface and the camera [72, 73, 80, 97, 75, 98, 59, 95]. Whereas some authors also calibrated the thickness of the port [72, 73, 84], others directly measured it or assumed it to be known [80, 75, 59, 95].

The errors in 2D imaging or 3D reconstruction with inaccurate measurements or inaccurate assumptions of the port's pose and thickness have not been investigated explicitly in prior research, yet. However, this knowledge is necessary to understand, which system parameters need to be modelled, calibrated or measured, respectively. It also indicates how precisely the camera and port need to be mounted. In this section, we explicitly investigate the required accuracies in port orientation, distance and thickness.

6.4.2. Proposed evaluation

In our proposed evaluation, we evaluate the imaging and reconstruction errors in a simulation, which is based on our proposed refractive forward projection method (chapter 5), our proposed stereo reconstruction method (section 6.2), our proposed dominating green wavelength (section 6.3), our Bumblebee stereo camera (Table 2.1) and the refractive indices, which are described in more detail in section 6.5. The port is placed at a front parallel pose 2 cm in front of the camera.

Table 6.1.: Maximum 2D and 3D errors at a distance of 1 m caused by inaccurate port modelling

	Range	Reference	2D [px]	3D [mm]
Yaw [°]	-1 - 1	0	18.2, 18.8	37.3, 37.1
Pitch [°]	-1 - 1	0	12.9, 13.9	11, 11.3
Distance [mm]	19 - 21	20	1.1, 1.1	2.1, 2.1
Thickness [mm]	11 - 13	12	0.2, 0.2	0.2, 0.2

To generate imaging and reconstruction errors, we now explicitly changed the translation, orientation and thickness of the port and calculate the arising projection and reconstruction errors. The orientation is parameterised by the yaw and pitch angle of the port assuming a horizontal stereo rig. To give an idea of how rigidly the port needs to be mounted and how accurately its thickness needs to be known, we considered for the yaw and pitch angle a variation of $\pm 1\,^\circ$ and for the port distance and thickness a range of ± 1 mm.

For the evaluation of the maximum imaging and reconstruction error, we consider the object point, which is located at a distance of 1 m, is visible in both camera images and exhibits the maximum angle to the optical axis. For our wide-angle underwater camera, the maximum angle in air is $56.5\,^\circ$ corresponding to the upper left corner of the right camera image. We also consider the upper right corner of the left camera image due to the asymmetry in 3D reconstruction if the port is rotated around the yaw axis.

6.4.3. Results

The results are shown in Table 6.1. With respect to the value ranges chosen, the projection and reconstruction errors due to an error in rotation are significantly high and amount in average to about ± 18.5 px and ± 37 mm for a rotation around the yaw axes and to about ± 13.4 px and ± 11.2 mm for a rotation around the pitch axis. The error caused by an inaccurate port distance, is considerably smaller (± 1.1 px, ± 2.1 mm) but still large enough to be measured. Our experiment also revealed that already small changes in the thickness of the port have measurable effects on projection and reconstruction. If the distance to the camera stays constant but the thickness of it is changed, errors of ± 0.2 px and ± 0.2 mm arise. These errors still increase, if only a thin port with no thickness is considered. In this case, the errors amount to considerable 2.4 px or 2.8 mm, respectively. These results clearly show that the port thickness should not be neglected in highly accurate, wide-angle, and high resolution underwater photogrammetry.

6.5. Effects of the index of refraction on projection and reconstruction

6.5.1. Related literature

Prior research does not give a clear picture on the relevance of the indices of refraction for accurate projection and reconstruction with flat port underwater (stereo) cameras or with in-air cameras imaging an object through the water surface, respectively.

In 1977, McNeil [99] modelled the index of refraction of water in order to preclude degradation in resolution caused by inaccurate focusing. McNeil [99] especially analysed how the index of refraction changes with respect to the wavelength of light, and the salinity, temperature and pressure (depth) of water, but did not analyse the projection or reconstruction errors. Treibitz et al. [68, 71] and Sedlazeck and Koch [100, 75] used the well-known value of $n_w = 4/3$ for the refractive index of water without explicitly considering the effects on projection and reconstruction. The decision to use this constant index of refractive of water was made by the two groups due to the 2-3 % variation of n_w with respect to the wavelength of light and the salinity, temperature and pressure of water, which appeared negligible. But this is not negligible, as our investigations will show. On the other hand, Shortis and Harvey [101] and Li et al. [95] mentioned that increased accuracy of their underwater photogrammetric systems is to be expected if the refractive index of water is incorporated in the calibration process. The expected improvement in accuracy was not investigated, though. Telem and Filin [98] explicitly calibrate the index of refraction of water, and Yau et al. [72] computed the refractive index based on the measurement of the dominating wavelength.

In our proposed work, we explicitly analyse the effects of the index of refraction of water on projection and reconstruction accuracy using the index of refraction of water model of McNeil [99]. Building on prior research, we not only evaluate the effects of the index of refraction of water, but also of the port (PMMA) and of air. This evaluation is accompanied by a study of the natural variations of the indices of refraction of water, PMMA and air.

6.5.2. Proposed evaluation of variations in the indices of refraction

6.5.2.1. The index of refraction of water

The index of refraction of water depends on the wavelength of light and on the salinity, temperature and pressure of water [99]. In order to determine the effect of these parameters on the projection and reconstruction accuracy, we firstly studied their natural variation based on physical oceanography literature of Reddy [102] and Talley et al. [103]. The effect on projection and reconstruction is subsequently evaluated in section 6.5.3.2.

6.5.2.1.1. Salinity Surface salinity varies with geographic location. In open oceans, it ranges from 33 ‰ to 37 ‰ and averages 34.6 ‰. Surface salinity mainly changes due to evaporation, precipitation or fresh water inflow. In more separated oceans with high evaporation, such as the Mediterranean or the Red Sea, surface water reaches high salinity values up to 39 ‰ or 41 ‰, respectively [103]. Surface waters with fresh water inputs and low salinity are located on coastlines and polar regions where fresh water is supplied by rivers or melting ice. In terms of latitude, maximum surface salinity is reached in the subtropics where evaporation dominates.

Water salinity also varies with depth and is more complex than the vertical distribution of water temperature. In contrast to water temperature, salinity has no dominating effect on the water density in general, and thus does not result in a monotonic vertical increase with depth. Exceptions are the polar and subpolar region, where the upper water layer may be coldest and salinity dominates water density and increases with depth. Here, a permanent rapid increase in salinity, called halocline, at depths from 25 to 100 m is found. The vertical change in salinity of the salinity profiles presented in [102, 103] is less than 2 ‰ for tropical, subtropical and subpolar regions or less than 4 ‰ for polar regions, respectively.

Surface water salinity is also subject to temporal changes. Seasonal changes in water salinity are mainly caused by the seasonal changes of evaporation, precipitation and fresh water inflow. In the open ocean, seasonal variations in salinity reach values of 0.5 ‰ and in the coastal waters of the Arabian Sea up to 4.6 ‰ [102]. Daily (diurnal) changes in surface water salinity are relatively small and do not exceed 0.21 ‰ in extreme cases.

6.5.2.1.2. Temperature Sea water surface temperature varies world-wide between −2 and 30 °C from the poles to the equator [102]. Seasonal temperature variations of open ocean surface water range from 1 to 5 °C at the equator and poles, and from 5 to 10 °C at mid latitudes [103]. Strong seasonal variations of 10 to 20 °C occur in coastal,

sheltered areas and in the western subtropics of the northern hemisphere at the Gulf and Kuroshi Stream [103]. Daytime temperature changes in the open ocean vary from 0.3 to 0.5 °C and between 2 and 3 °C in shallow and sheltered coastal regions [102]. Rapid vertical declines in temperature, so called thermoclines, are particularly present at low and middle latitudes and are close to the surface in summer. Here, typical temperature declines from 20 to 8 °C from the water surface down to 500 m are observed [102, 103]. Even stronger temperature decreases, so-called metalimnions, occur in lakes in warm seasons. Within a narrow range, the temperature rapidly drops from the surface temperature to about 4 °C at the bottom [104].

6.5.2.1.3. Pressure The deepest known point of all oceans is located in the Mariana Trench at a depth of 10.911 km. The major part of the ocean (99.85 %) is located between 0 and 7000 m [99]. The water pressure P increases with the depth Z roughly by 1 dbar/m. More precisely, the water pressure depends also on gravity and water density, which are a function of the latitude ϕ or the vertical water temperature profile $T(Z)$ and the vertical salinity profile $S(Z)$, respectively. In 1983, an accurate model $f_Z^* : (P, \phi, T(Z), S(Z)) \mapsto Z$ to calculate the depth in sea water was described by the UNESCO formula of Fofonoff and Millard [14]. If the exact vertical salinity and temperature profile, $S(Z)$ and $T(Z)$ are not known, the maximum error of $f_Z : (P, \phi) \mapsto Z$ assuming standard ocean water ($S = 35$ ‰, $T = 0$ °C) amounts to 2.1 m at 10.000 dBar.

To evaluate the water pressure as a function of water depth $f_P : (Z, \phi) \mapsto P$, we calculated the inverse of f_Z using Newton's method with 3 iterations. Our proposed model reads as follows:

$$\begin{aligned} f_P : (Z, \phi) &\mapsto P_3, \\ P_{i+1} &= P_i - \frac{e_i}{e_i'},\ P_0 = Z \end{aligned} \tag{6.9}$$

where Z is given in metres, ϕ in degrees, and P in decibars and where

$$\begin{aligned} e_i &= (c_1 P_i + c_2 P_i^2 + c_3 P_i^3 + c_4 P_i^4)/(g + c_5 P_i) - Z, \\ e_i' &= (c_1 g + 2c_2 g P_i + (c_2 c_5 + 3c_3 g) P_i^2, \\ &\quad + (2c_3 c_5 + 4c_4 g) P_i^3 + 3c_4 c_5 P_i^4)/(g + c_5 P_i)^2, \\ g &= c_6 (1.0 + c_7 \sin^2 \phi + c_8 \sin^4 \phi), \end{aligned}$$

$$
\begin{aligned}
&c_1{=}9.72659, c_2 = -2.2512 \cdot 10^{-5}, c_3 = 2.279 \cdot 10^{-10},\\
&c_4{=}-1.82 \cdot 10^{-15}, c_5 = 1.092 \cdot 10^{-6}, c_6 = 9.780318,\\
&c_7{=}5.2788 \cdot 10^{-3}, c_8 = 2.36 \cdot 10^{-5}.
\end{aligned} \tag{6.10}
$$

Using this proposed model, the water pressure at 7000 m is ∼7155 dbar at the equator and increases to ∼7193 dbar at the poles.

6.5.2.2. The index of refraction of PMMA

The port of our underwater housing is made out of optical-grade cast acrylic plastic. Acrylic plastic is PMMA and is a widely used optical plastic whose index of refraction is also wavelength and temperature dependent.

To model the index of refraction of PMMA as a function of the wavelength and of the temperature, we used the models of Kasarova et al. [105] and of Cariou et al. [106]. Kasarova et al.'s model $f_{n_p,\lambda} : \lambda \mapsto n_p$ describes the wavelength dependency of the index of refraction of PMMA at constant temperature of 20 °C. Cariou et al. modelled the variation dn_p/dT of the refractive index with respect to the material's temperature T at constant wavelength. To determine the the index of refraction of PMMA with respect to our temperature ranges and specific wavelengths, we combined both models assuming no correlation between λ and T. The combined proposed model looks as follows:

$$
\begin{aligned}
&f_{n_p} : (\lambda, T) \mapsto n_p,\\
&n_p{=}f_{n_p,\lambda}(\lambda) - 1.258 \cdot 10^{-4}(T - 20),
\end{aligned} \tag{6.11}
$$

where T is valid for the temperature range of -23 to 93 °C. For the accurate formula of $f_{n_p,\lambda}(\lambda)$, we refer to [105].

6.5.2.3. The index of refraction of air

Several comprehensive models of the index of refraction of air are available. We used the model of Ciddor [107], which describes the index of refraction of air as a function of wavelength, temperature, pressure, humidity and CO_2 concentration. Whereas the temperature of air may vary with the water temperature during a mission, we assume the pressure, humidity and CO_2 concentration to be constant after the underwater housing was closed. Assuming a solid, non-deformable housing, the average air pressure depends on the altitude above sea-level when the housing was closed. A description of standard

air pressure as a function of the altitude above sea-level is part of the model collection of the standard atmosphere established by COESA [108].

6.5.3. Results

To determine the maximum projection and reconstruction errors due to inaccuracies in the dominating wavelength and inaccuracies in the refractive indices of water, PMMA and air, a similar simulated experimental setup as in section 6.4.2 together with the proposed index of refraction models is used.

6.5.3.1. Errors by neglecting the wavelength of light

As described in section 6.5.2.1, 6.5.2.2, and 6.5.2.3, water, PMMA and air are subject to dispersion. Based on the three domination wavelength $\lambda_r = 630\,\text{nm}$, $\lambda_g = 538\,\text{nm}$ and $\lambda_b = 463\,\text{nm}$ of the camera sensors (section 6.3) and using λ_g as reference, we obtained a projection error of 4.3 px or 5.5 px if light with a different wavelength λ_r or λ_b, respectively, was projected. The corresponding reconstruction errors amounted to 5.4 mm (λ_r) and 6.9 mm (λ_b). The results show that wavelength dependent effects in flat refractive geometry are clearly visible and that the dispersion of water mainly contributes to these effects. In real images, this so called chromatic aberration is recognisable as colour seams or blurry edges. Our analysis also showed that dispersion significantly deteriorates accuracy in photogrammetry and should not be neglected. Due to the large number of wavelengths in the sensed light spectrum, chromatic aberration and wavelength dependent reconstruction errors are the only errors, which cannot be removed completely by correcting the indices of refraction. In section section 10.2, we propose an approach to partially minimise chromatic aberration by warping the colour channels according to their dominating wavelength. In our proposed refractive reconstruction method in chapter 8, we reduced the wavelength dependent reconstruction errors by considering only a single colour channel.

6.5.3.2. Errors by an inaccurate refractive index of water

As studied in section 6.5.2.1, the index of refraction is also a function of the salinity, temperature and pressure of water.

To evaluate the projection and reconstruction errors caused by variations in salinity, we considered the case that the camera was at first calibrated in a fresh water tank with no salinity and afterwards used in the very salty water of the Red Sea (41 ‰). In this

Table 6.2.: The indices of refraction and the resulting maximum projection error (2D) and maximum reconstruction error (3D) at a distance of 1 m caused by changes in the parameters of light, water, PMMA and air

	Range	Reference	Model	Refractive index μ	2D [px]	3D [mm]
Water			[99]			
wavelength [nm]	630-463	538		1.3318-1.3347-1.3384	4.3, 5.5	5.4, 6.9
salinity [‰]	0-41	0		1.3347-1.3423	11.3	14
temperature [°C]	-2-30	20		1.3357-1.3335	1.5, 1.8	1.8, 2.3
depth (pressure) [m]	0-7000	0	[14]	1.3445	14.9	18.3
PMMA (acrylic)						
wavelength [nm]	630-463	538	[105]	1.4891-1.4943-1.4989	0.06, 0.05	<0.1
temperature [°C]	-2-30	20	[106]	1.4969-1.4931	0.03, 0.02	<0.1
Air			[107]			
wavelength [nm]	630-463	538		1.000271-1.000273-1.000275	<0.01	<0.1
temperature [°C]	-2-30	20		1.00029-1.00026	0.05, 0.02	<0.1
altitude (pressure) [m]	0-8848	0	[108]	1.00027-1.00008	0.37	0.5
humidity [%]	0-100	85		1.000272-1.000273	<0.01	<0.1
CO_2 [ppm]	388-408	398		1.00027259-1.00027260	<0.01	<0.1

case, our simulation reveals a significant projection and reconstruction error of 11.3 px and 14 mm, respectively. Also smaller changes in salinity due to changes in depth are measurable. The halocline at high latitude from 33.7 ‰ at the water surface up to 34.9 ‰ at a depth of 42 m as described by Reddy [102] results in a projection error of 0.4 px (0.4 mm), which is measurable in computer vision, where image coordinates, in general, can be determined up to an accuracy of ∼0.1 px.

Projection errors are also measurable under changing water temperature. Natural ocean water temperature ranges from −2 to 30 °C. Assuming, for instance, a camera calibrated in a water tank at a water temperature of 20 °C, the resulting errors amount to 1.5 px and 1.8 mm or 1.8 px and 2.3 mm, if the camera was subsequently used at a temperature of −2 or 30 °C, respectively. Relatively large errors also occur due the rapid temperature declines in lakes. Assuming a camera calibrated at low depth at 25 °C, the errors at the bottom with 4 °C reach 2.3 px or 2.8 mm, respectively.

The water pressure is also a significant variable in accurate imaging and photogrammetry. It especially becomes important at high depth. If a camera is calibrated at sea-level, for instance, and then used at a depth of 7000 m, the projection and reconstruction error amount to significant 14.9 px and 18.3 m. Herewith, variations in pressure cause the largest errors in this error analysis of the index of refraction of water.

6.5.3.3. Errors by an inaccurate refractive index of PMMA

The projection and reconstruction error were also measured under temperature changes of the port, which is made out of PMMA. Assuming a calibration at a reference temperature of 20 °C, the projection errors at 0 or 30 °C amount to 0.03 px or 0.02 px, respectively, and are not measurable with our image resolution. Also the reconstruction error is small and less than 0.1 mm.

6.5.3.4. Errors by an inaccurate refractive index of air

The inaccuracy of the underwater camera was also investigated under different temperature, pressure, humidity and CO_2 concentration of air. As our simulations showed, the errors caused by humidity (0 to 100 %) and CO_2 concentration (388-408 ppm) are negligible and less than 0.01 px or 0.1 mm. The temperature effect is stronger but still very small. With respect to the reference temperature of 20 °C, the projections errors at −2 or 30 °C amount to 0.05 or 0.02 px, and the reconstruction errors are less than 0.1 mm. The pressure of air is the only parameter whose influence can be actually measured. Assuming that the camera housing was sealed under the extreme low pressure conditions

on the peak of Mount Everest (8848 m) rather than under the higher pressure at sea-level, measurable errors of 0.37 px and 0.5 mm could be simulated. A projection error of 0.1 px is even measurable if the housing was closed at a lower altitude of ∼1600 m.

6.6. Conclusion

Our evaluation of the indices of refraction showed that underwater cameras with a flat port are subject to strong chromatic aberration. This effect is always visible and should not be neglected in accurate, wide-angle refractive imaging and photogrammetry. Among the three indices of refraction, the index of refraction of water has the most significant effect on the projection and reconstruction behaviour of the optical system. Considering only the natural value ranges of the parameters of the index of refraction of water, the pressure of water has the strongest effect on projection and reconstruction accuracy (14.9 px, 18.3 mm), followed by salinity (11.3 px, 14 mm), wavelength (9.8 px, 12.3 mm) and temperature (2.3 px, 4.1 mm). Our evaluation also revealed that significant changes in refractive geometry are not only expected between different geographical locations, but also during underwater missions. Changes in operating water depth might cause significant changes in pressure, salinity and temperature and thus deteriorate the system's measurement accuracy.

Depending on the field of application and the accuracy required, we recommend the re-adjustment of the index of refraction of water, if the geographical location has changed or strong changes in pressure, salinity or temperature occur during a mission. In contrast to existing literature, we hypothesise that the calibration before a mission of the index of refraction of water or even of the whole camera housing with a calibration target is time consuming and not as accurate as a calibration in a water tank under optimal conditions. For reduced effort and higher accuracy, we therefore recommend only a one-time water tank calibration in combination with a subsequent indirect on-site measurement of the index of refraction with a salinometer, thermometer and pressure sensor or an approximate estimate of the water depth for shallow water applications, respectively. An underwater camera equipped with theses sensors has also the advantage of online calibration during a mission to compensate for the effects of varying pressure, salinity and temperature.

In addition to the index of refraction of water, we investigated the effects of the index of refraction of air and PMMA (port). Our simulations showed that the air pressure under which the underwater housing was sealed has measurable effects. Our analysis also

showed that the impact of temperature on the indices of refraction of PMMA and air is relatively small and not measurable with our ∼0.8 MP wide-angle underwater camera. If a image sensor with a higher resolution is used, this may no longer be negligible.

We also evaluated the projection and reconstruction errors if the pose and thickness of the port are not exactly known. Those inaccuracies might be caused by a not completely rigid mounting of the camera or of the port, by inaccurate measurement of the thickness of the port or by single interface approximation. Our experiments revealed that the orientation and especially the rotation around the yaw axis led to the largest errors measured in this chapter and should not be neglected. Our work also showed that the effects of the port distance and thickness are significantly smaller but should not be neglected either. In contrast to prevailing literature, we demonstrated the importance of accurate port thickness modelling using high resolution, wide-angle underwater stereo cameras embedded in a flat port underwater housing.

7. Accurate and time-efficient calibration of underwater cameras with a flat thick port

7.1. Introduction

Until today, the physics-based refractive model for flat port underwater cameras has not attracted much attention. Practically all well-established underwater vision research projects on underwater reconstruction, localisation or mapping, such as [20, 49] [109, 110] [111, 51] [112, 113] [50], still use the standard perspective camera model and corresponding calibration techniques [114] to compensate for the effects of refraction. However, the perspective camera model with a modified focal length in combination with the standard lens distortion model cannot imitate the 3D distortion, which depends on the object distance [59]. The alternative physics-based underwater camera model used in our proposed underwater camera calibration method explicitly incorporates the refraction of light at the air-glass and glass-water interfaces and models the optical system more precisely.

Flat port underwater camera calibration based on the physics-based refractive camera model is still an active field of research. Surprisingly, no calibration method has been presented yet, which is similar to standard in air camera calibration techniques, where multiple views of a calibration target are used, and the parameters are estimated by minimising the reprojection error. Moreover, almost no prior literature elaborated on the importance of the wavelength of light (chapter 6) or even tried to calibrate for this.

In this chapter, we propose a calibration framework for underwater stereo cameras with a thick flat port, which is based on the physics-based refractive underwater camera model and calibrates the orientation and distance of the port relative to the camera, and the dominating red and blue wavelengths. In contrast to prior literature, the parameters are calibrated from multiple images by minimising the reprojection error using our previously

proposed efficient and accurate refractive forward projection method (chapter 5). Based on our experimental results in a water tank, we recommend the additional calibration of the baseline of the stereo camera and the calibration with images from different camera views.

7.2. Related literature

Numerous studies on the calibration of underwater cameras exist. However, we limit our focus here to literature on underwater camera calibration, which is based on the physics-based flat port underwater camera model.

In 1988, Kotowski [84] explained, how ray tracing, refractive forward projection and bundle adjustment in combination with numerical optimisation can be used to determine the position and shape of multiple refractive interfaces and the relative indices of refraction. However, no practical results were presented.

In 1997, Li et al. [95] reported on the calibration of stereo cameras embedded in two separate underwater housings with a flat port and a baseline of 40 cm. The calibration setup consisted of a relatively large three-dimensional calibration frame of $1.4 \times 1.4 \times 0.7\,\mathrm{m}^3$. The calibration was performed in a two-step approach with an empty and subsequently filled water tank. In the first step, the cameras' intrinsic and extrinsic parameters were calibrated. In the second step, the tank was filled with water to calibrate the poses of the ports. The thickness of the port and the indices of refraction of water and glass were not calibrated but measured beforehand. For the calibration, one stereo image pair with 15 manually determined control points was taken to calculate the 2×3 port pose parameters ($1\times$ distance, $2\times$ orientation). The parameters were obtained by minimising the 3D reconstruction errors using refractive back projection.

Due to the physics-based camera model, the underwater camera calibration method of Li et al. was relatively advanced at the end of the 1990s but still limited by the available computational power. This reflected in the usage of only one image pair, the low number of control points (corners), and the manual corner detection. The two step calibration with and without a filled water tank and the large calibration frame made the method also very time consuming. In contrast to standard in-air calibration techniques, Li et al. minimised the reconstruction error instead of the reprojection error. Minimising the reprojection error would have been more accurate, as the noise in the image space can be accurately modelled by normal distribution. Like the standard calibration method for the calibration of cameras in air, our proposed calibration method

minimises the reprojection error, uses a more convenient flat calibration checkerboard and can process hundreds of image pairs with a much higher number of control points in a short calibration time of only few seconds.

In 2008 and 2012, Treibitz et al. [68, 71] described the calibration of a short range measurement system (∼0.5-1.5 m) consisting of a monocular underwater camera with in-situ adjustable zoom and focus. The thickness of the port was simplified by single interface approximation and the orientation of the port assumed parallel to the camera's image plane. A flat calibration target with known point distances (checkerboard) was used and had to be placed at a known distance parallel to the camera's image plane. For the calibration of the port distance and focal length, the distance errors to the object points of the calibration target were minimised. Minimising the total reprojection error was avoided, as there was to be no closed-form solution, which was allegedly required. Due to its variation of less than 3 %, the index of refraction of water was not explicitly calibrated, but assumed to be known ($n_w = 4/3$).

Treibitz et al. claimed that there is no closed-form solution for the refractive forward projection with a single interface. As shown by Glaeser and Schröcker [83] and described in chapter 5, this is not correct, since the refractive forward projection with a single interface can be formulated by a polynomial of degree four (quartic equation) for which a closed-form solutions exists [85]. Moreover, a closed-form solution for the refractive forward projection is not explicitly required but can be determined numerically (see chapter 5). The calibration method of Treibitz et al. contains numerous simplifications or inaccuracies, and thus is only appropriate for close-range size measurements with accuracies in the order of centimetres. The assumption of parallelism between the port and the camera image plane is critical. As shown in chapter 6 for a short baseline wide-angle underwater stereo camera, errors in the orientation of the port can lead to large reconstruction errors. Another inaccuracy arises by using a constant index of refraction of water. As also shown in chapter 6, small changes of the index of refraction of water lead to significant reconstruction errors. Furthermore, the requirement to measure the object distance is inconvenient and represents another source of error. Similar to Li et al. [95], the cost function to be minimised is defined in the object space instead of the image space (reprojection error) and thus further increases the total measurement error.

In contrast to the calibration method of Treibitz et al., our proposed method uses a more accurate underwater camera model with arbitrary port orientation and port thickness. Furthermore, our method minimises the reprojection error and neither requires

parallelism between the checkerboard and the camera image plane nor the measurement of the checkerboard distance.

In their paper from 2008, Kunz and Singh [59] described the theoretical calibration of a monocular flat port camera, but did not present any results. In the described calibration method with checkerboard images from one or multiple camera views, the authors estimated the pose of the port (3 parameters) and the relative pose between the camera and the n checkerboards ($n \times 6$ parameters) by minimising the reprojection error using refractive forward projection.

We actually implemented this theoretical idea for a stereo camera, and show that the calibration with multiple images using refractive forward projection leads to a stable and accurate convergence behaviour and can be carried out with hundreds of images in only few seconds.

In their calibration method for monocular underwater cameras from 2010, Telem and Filin [98] applied single interface approximation and approximated refractive forward projection. The refractive forward projection for each projected image point was imitated by a perspective camera, whose focal length imitates the projection on the refractive interface by perspective projection. Using virtual perspective cameras enabled the calibration of the port distance, the port orientation and the index of refraction of water with standard in-air calibration techniques. In Telem and Filin's approach, the pseudo reprojection error was calculated on the refractive interface. The accuracy and convergence behaviour of the calibration method was studied in a simulation with various calibration targets (2D planar, 3D lattice), image numbers (1-16), camera view distributions, levels of pixel noise and initial guesses. All setups showed stable convergence behaviour and allowed an accurate computation of the index of refraction of water. For accuracy and convenience reasons, and even though the calibration was already possible with a single calibration image, the authors recommended a calibration with a planar checkerboard with multiple images from different camera views. The calibration accuracy and the multi-view reconstruction accuracy were also evaluated in a swimming pool and the open sea. The calibration with 6 images of a calibration plate and additional objects of known size located at an object distance of $\sim$1 m led to a standard deviation of 0.4 px. The reconstruction accuracy was evaluated by considering the distance between known points or by considering the distances of a set of points to a line or a plane.

In contrast to Telem and Filin, we were not able to calibrate the index of refraction of water or the dominating wavelength of the green colour channel, respectively, due to correlations with the port distance and the distance of the calibration target. Instead,

we use the dominating green wavelength of the green colour channel from the data sheet of the camera sensor and calibrate in addition to that the dominating red and blue wavelengths. Telem and Filin do not consider wavelength dependent refraction at all. Furthermore, we model the port thickness and minimise the actual 2D reprojection error. As we have shown in chapter 9, single interface approximation leads to measurable imaging and reconstruction errors and is not necessary for time-efficiency reasons (see chapter 5). Using the reprojection error is in general more accurate, as the error distribution can be modelled more accurately by bivariate normal distribution. Telem and Filin recommended multi-view calibration images based on experiments with simulated data, our practical water tank experiment confirms this recommendation. Although our experiments with simulated data worked with a single camera view, the experiments with real data revealed the limitations of the physics-based model and the limitations of the standard corner detection method. To compensate for these inaccuracies, we recommend the recalibration of the baseline of the stereo camera with multiple images from different views. In this way, we could halve the root mean square (RMS) residual error to only $\sim$0.19 px.

In 2011, Gedge et al. [70] calibrated the orientation and distance of the port of an underwater stereo camera. The model was simplified by single interface approximation. The calibration data consisted of the image points of arbitrary objects points of the scene, which were extracted from a calibration checkerboard in the experiments. The pose of the port was estimated by nonlinear optimisation in which the inverse depth-weighted distance between the back-projected rays of the stereo correspondences was minimised. Refractive forward projection with quartic equation-based refractive forward projection (see chapter 5) was avoided due to efficiency reasons.

In contrast to Gedge et al., our calibration is more accurate, as we also model the port thickness, consider the geometric relations between the object points and minimise the reprojection error. We also show that refractive forward projection does not represent a bottleneck in flat port underwater camera calibration.

In 2011, Sedlazeck and Koch [75] presented the calibration of a stereo camera with two separate underwater housings. Due to an allegedly extremely time-consuming computation of the reprojection error, the point-wise perspective camera method of Telem and Filin [98] was used. Sedlazeck and Koch extended Telem and Filin's method for thick ports and also modified the virtual centre of projection of the camera. Erroneously, the virtual centre of each camera was displaced from the originally correct virtual centres of projection [98] to a point on the envelope of the refracted light rays, also referred

to as caustic [76]. The correct virtual centre of projection is actually located on the axis of the axial camera [73]. Sedlazeck and Koch decided to calibrate the port distance and orientation of each camera and the relative pose between the two cameras without a checkerboard, but by arbitrary objects points of the environment. The implemented bundle adjustment approach, in which additionally the pose of the stereo camera and the position of the object points were calibrated, minimised the reprojection error on the port-water interface, but was very time-consuming. For one image pair, the implemented algorithm required about 5 minutes and for multiple image pairs several hours. The accuracy of the calibration method was evaluated by synthetically generated image points and rendered images in combination with certain feature matching techniques. Real images were only used for evaluating the robustness of the calibration of the ports' orientations.

This approach has some disadvantages due to the extremely long calibration time, the limited accuracy, the missing evaluation of the method under real conditions, and the missing incorporation of chromatic aberration in image rendering. Even though our method is computationally less demanding, as we use a stereo camera with a single underwater housing and a calibration checkerboard, our proposed method is significantly faster and requires only $\sim$0.9 seconds for 100 image pairs. Sedlazeck and Koch's calibration method is also more inaccurate, as Telem and Filin's [98] approximated refractive forward projection was used, which was additionally modified incorrectly. There is only little information on the accuracy of the calibration method under real conditions. In contrast to that, we evaluate in this section the reprojection error, in chapter 8 the accuracy in 3D reconstruction and in chapter 10 the capability to minimise chromatic aberration.

In another paper of Jordt-Sedlazeck and Koch [80] from 2012 on the calibration of underwater stereo cameras with a flat port and separate underwater housings, the authors applied analysis-by-synthesis and evolutionary optimisation. In this calibration approach, a calibration checkerboard was used whose images were back-projected onto the estimated checkerboard plane and which were there compared with rendered checkerboard images. This analysis-by-synthesis approach was preferred to classic checkerboard calibration methods so as to become independent of errors in corner detection. The difference in greyscale values between the pixels of the back-projected and rendered images were minimised by evolutionary optimisation, which was favoured over derivative-based optimisation to avoid getting stuck in local minima.

As mentioned by the authors, evolutionary optimisation is very time-consuming. It is

unclear if evolutionary optimisation or analysis-by-synthesis are necessary at all. Jordt-Sedlazeck and Koch did not give any example in which derivative-based optimisation or classic corner detection fail. Our derivative-based calibration algorithm converged in all cases. Furthermore, the effect of wavelength dependent light refraction, which is visible as chromatic aberration, was not considered in the rendering process. In contrast to that, we explicitly incorporate chromatic aberration in the calibration process by calibrating the dominating wavelength of the red and blue colour channel.

In a recent publication from 2012, Agrawal et al. [73] investigated the calibration of one or multiple parallel flat refractive interface with a monocular camera, which also included the theoretical analysis of underwater cameras with a flat, thick port. Agrawal et al. used a single calibration image of a checkerboard to estimate the camera-relative poses of the interfaces and of the checkerboard. The calibration consisted of the computation of an initial guess of these parameters using a set of linear constraints, and a subsequent non-linear refinement. In the first step of the initial guess computation, the axis of the axial camera, the relative camera-checkerboard rotation and the checkerboard translation lateral to the axis of the axial camera were computed. These parameters are independent of the locations of the interfaces and of the indices of refraction and thus can be also obtained in-air. Here, the fact was exploited that each refractive forward projection takes place in the plane of incidence, which always contains the axis of the axial camera and the corresponding image and object points. In the second step of the initial guess computation, the camera relative distances of the interfaces and of the checkerboard were determined. This step depends on the indices of refraction and needs to be conducted in water. The final non-linear refinement was demonstrated for the air-water case (single interface) and the air-water-air case (imaging through a water tank). For these cases, Agrawal et al. demonstrated for the first time a working calibration system, which minimised the reprojection error. Here, refractive forward projection was realised by solving a polynomial of degree 4.

The work of Agrawal et al. is one of the first publications on flat port underwater camera calibration in which the reprojection error is minimised. However, only the computationally less expensive cases with 4th degree polynomial refractive forward projection and with only one calibration image were evaluated in practice. Similar to Agrawal et al.'s work, we also use refractive forward projection to minimise the reprojection error. But we propose the calibration of underwater cameras with a thick port, which is computationally more demanding. To increase the accuracy, we additionally use multiple images and also evaluate the calibration time. Even though our system is

computationally more expensive, as it takes into account the thickness of the port and uses a high number of calibration images, we show that the calibration is still efficient.

In 2012, Kang et al. [82] reported on a two-view underwater structure and motion framework with two separate flat port underwater cameras. Here, the camera poses, the positions of the object points and the distances of the ports were estimated. The orientations of the ports were assumed to be perfectly parallel to the cameras' image planes. The calibration data was extracted from the observed scene and did not require any additional calibration target. According to the author's, the scale of the setup cannot be recovered and thus the port distance can be only estimated with respect to a certain scale factor. In the last step of the estimation, the parameters were determined by minimising the reprojection error in a bundle adjustment framework.

Kang et al. assumed parallelism between the port and the camera. As shown in chapter 6, this leads to large imaging and reconstruction errors. Moreover, recovering the scale of the scene with monocular flat port underwater cameras is possible in principle, as those cameras do not exhibit a perspective imaging behaviour. This was shown by Jordt-Sedlazeck and Koch [79] at least with synthetic noise-free data. In contrast to Kang et al.'s work, we also calibrate the orientation of the port and also model the thickness of the port to achieve a higher accuracy in 2D imaging and 3D reconstruction.

The work of Yau et al. [72] from 2013 is inspired by the work of Agrawal et al. [73] and reports on the calibration of one or multiple parallel refractive interfaces with a monocular camera and deploys a special calibration target with light points that emit light of two different wavelengths. To improve the accuracy of the initial guess for the subsequent nonlinear refinement step, Yau et al. [72] incorporated additional constraints based on the wavelength dependent refraction of light. Firstly, the initial guess of the axis of the axial camera was calculated by just considering the image point-pairs created by the projections of an object point with light of two different wavelength. These image point-pairs always lie in the same plane of incidence. Secondly, to calculate the initial guess of the interface distances and the initial pose of the calibration target, Yau et al. extended Agrawal et al.'s method by incorporating the wavelength dependent triangulation between each image point-pair and the corresponding object point. In the final non-linear refinement step, the total reprojection error was minimised, which was obtained by back projection-based refractive forward projection. As one of the first authors, Yau et al. demonstrate refractive forward projection-based calibration with two refractive interfaces in practice. Yau et al. compared their calibration method with the work of Agrawal et al. using simulated and real data. The method of Yau et al. showed

in all simulated cases (single interface, thick port with/without thickness estimation) with one calibration image significantly more accurate initial guesses. The more accurate initial guesses also reflected in the better results after non-linear refinement and particularly improved the thickness estimation of the port. The calibration method was also evaluated under real conditions with a colour camera placed in front of a water tank in which the light-emitting diode (LED) calibration target was submerged. The lateral chromatic aberration of the camera lens was minimised by colour channel specific calibration of the lens distortion coefficients. Before the red and blue LED points were extracted from the image data, the red and blue colour channel were split from the raw image according to the Bayer colour filter array and subsequently demosaiced.

Similar to our proposed calibration method and as one of the first authors, Yau et al. explicitly considered the wavelength dependency of the index of refraction (dispersion) and the minimisation of the reprojection error with two refractive interfaces. However, the calibration device deployed is cumbersome and harder to manufacture than an ordinary calibration checkerboard as used in our method. We also use more accurate models of the index of refraction of water and air. In addition to the wavelength of light and the temperature of water, we incorporate the salinity and pressure of water, which is necessary for the deployment in the sea, and do not assume an index of refraction of air of $n_a = 1$, which leads in our wide-angle underwater camera to an error of more than 0.4 px (see also chapter 6). To enable more accurate calibration with a checkerboard, we narrow the light spectrum down by using only the green colour channel for the port calibration, rather than using LEDs with distinct wavelengths. We do not think that exploiting multiple colour channels and wavelength triangulation will significantly contribute to the accuracy of our underwater stereo camera calibration method. Since the resolution of the red and blue colour channel in the Bayer colour filter array is lower than the resolution of the green colour channel (25 % versus 50 %), it makes more sense to use multiple green colour channel images as proposed in our method. In our stereo system, the depth information from wavelength triangulation is of little value due to the much larger baseline of the stereo camera. For simplicity, we omit the explicit correction of the lateral chromatic aberration of the lens, but calibrate the dominating red and blue wavelength with the physics-based flat port underwater camera model, which compensates indirectly also the chromatic aberration of the lens. In contrast to Yau et al. or Agrawal et al., the accurate computation of the initial guess is of minor importance to us. We use a straight forward initial guess algorithm, which exploits the depth information of the stereo camera. Although it might be less accurate than the

more sophisticated methods of Agrawal et al. and Yau et al., our proposed stereo camera calibration exhibits a robust convergence behaviour under noise and results in a several times better accuracy in the estimated port orientation and port distance. If we directly compare the precision of our (wide-angle underwater) stereo camera calibration with the monocular camera calibration results of Yau et al. after bundle adjustment using simulated noisy data ($\sigma = 0.5\,\mathrm{px}$), our approach shows an about 5.2 times lower port orientation error (axis error) and an about 3.5 times lower port distance error. Our work also elaborates on the calibration time and shows that accurate calibration of a thick port and hundreds of stereo calibration images is possible in only few seconds.

7.3. Proposed calibration method

Our proposed calibration method has been developed for an underwater camera consisting of a pre-calibrated stereo colour camera, which is embedded in a waterproof underwater housing with a thick flat port. The calibration procedure is almost identical to common in-air camera calibration where several images of a checkerboard are taken. The main difference to in-air camera calibration is that the calibration has to take place in water.

Our method calibrates the pose of the port $\boldsymbol{\pi}_p$, the poses $\boldsymbol{\pi}_{cb,1}, \ldots, \boldsymbol{\pi}_{cb,n}$ of the checkerboard and the dominating wavelengths, λ_r and λ_b, of the red and blue colour channel (see Fig. 7.1). The pose of the port $\boldsymbol{\pi}_p = (\theta_y, \theta_p, d)$ relative to the stereo camera is parameterised by the yaw and pitch angles, θ_y and θ_p, and the distance d. As a consequence of our water tank experiment, we later also calibrate the baseline b of the stereo camera.

Our calibration method requires a pre-calibrated stereo camera with rectified camera images and assumes the thickness of the port and the dominating wavelength of the green colour channel to be known. As a result of our research described in chapter 6, for higher accuracy, we also consider the salinity, the pressure and the temperature of water, and the pressure of air within the underwater housing.

Whereas the temperature of water can be easily measured with a thermometer, we used for the measurement of the salinity a salinometer, which determines the salinity indirectly by the water's conductivity. To determine the approximate water and air pressure, it is sufficient to know the approximate depth in water and the rough altitude in-air at which the calibration is carried out and at which the camera housing was sealed. The required water and air pressure can be then determined by the models described in

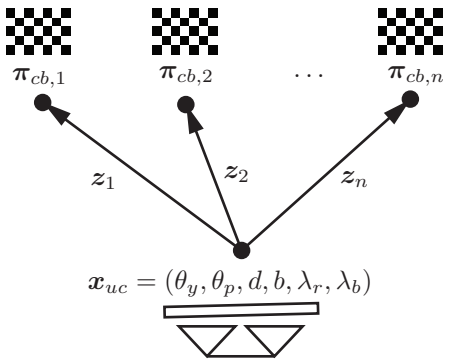

Figure 7.1.: Underwater camera calibration with multiple checkerboard image pairs: The figure illustrates the parameters involved. (θ_y, θ_p, d) denotes the pose of the port, b the baseline of the stereo camera and λ_r and λ_b the dominating wavelength of the red and blue colour channel. The calibration is based on the measurements $\boldsymbol{z}_1, \ldots \boldsymbol{z}_n$ and also requires the estimation of the relative camera-to-checkerboard poses $\boldsymbol{\pi}_{cb,1}, \ldots, \boldsymbol{\pi}_{cb,n}$.

chapter 6.

Our method is subdivided into several consecutive steps. Firstly, the initial guess of the pose of the port and of the poses of the checkerboard relative to the stereo camera are determined. Secondly, the pose of the port and the poses of the checkerboard, and thirdly the dominating wavelengths of the red and blue colour channel are refined by sparse bundle adjustment.

7.3.1. Initial guess of the port and checkerboard poses

A good initial guess $\boldsymbol{\pi}_{p,0}$, $\boldsymbol{\pi}_{cb,1,0}, \ldots, \boldsymbol{\pi}_{cb,n,0}$ of the pose of the port and of the poses of the checkerboard needs to be chosen to enable the optimisation algorithm in the bundle adjustment framework to converge against the globally optimal solution. Since the port is approximately parallel to the stereo camera ($\theta_y = \theta_p \approx 0$) and its approximate distance $\tilde{d}$ can be roughly measured or estimated, determining the initial pose $\boldsymbol{\pi}_{p,0} = (0, 0, \tilde{d})$ of the port is straightforward. The pose of each checkerboard is determined by Horn's absolute orientations algorithm [115], which allows the relative pose computation between two sets of 3D points with at least three given point correspondence. We use three points given in the checkerboard coordinate system and the three corresponding point reconstructions based on our initial underwater camera guess. The 3D point reconstruction used is explained in more detail in section 6.2.

7.3.2. Refining the port and checkerboard poses by sparse bundle adjustment

Our proposed underwater camera calibration uses sparse bundle adjustment and minimises the total reprojection error $e(\boldsymbol{x}, \boldsymbol{z})$ of all observed checkerboard corners $\boldsymbol{z}$

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} e(\boldsymbol{x}, \boldsymbol{z}), \tag{7.1}$$

to calibrate the pose of the port and the poses of the checkerboard $\boldsymbol{x} = (\boldsymbol{\pi}_p, \boldsymbol{\pi}_{cb,1}, \ldots, \boldsymbol{\pi}_{cb,n})$. The orientation of each checkerboard is represented by a quaternion. Using n stereo camera poses and checkerboards with n_c corners, the total reprojection error in the least-squares sense is defined as [116]

$$\begin{aligned} e: \ \mathbb{R}^{3+7n} \times \mathbb{R}^{4nn_{cb}} &\to \mathbb{R}, \\ (\boldsymbol{x}, \boldsymbol{z}) &\mapsto e(\boldsymbol{x}, \boldsymbol{z}), \\ e(\boldsymbol{x}, \boldsymbol{z}) &= \sum_{i=1}^{n} \sum_{j \in \{l,r\}} \sum_{k=1}^{n_{cb}} \boldsymbol{e}_{i,j,k,g}^T \boldsymbol{e}_{i,j,k,g}, \end{aligned} \tag{7.2}$$

where $\boldsymbol{e}_{i,j,k} \in \mathbb{R}^2$ denotes a single reprojection error.

For the calibration of the pose of the port and the poses of the checkerboard, we use the image points

$$\boldsymbol{z} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n\} = \{\boldsymbol{u}_{i,j,k} \in \mathbb{R}^2\} \tag{7.3}$$

from the green colour channel, where $i \in \{1, \ldots, n\}$ denotes the measurement index, $j \in \{l, r\}$ the left or right camera index and $k \in \{1, \ldots, n_{cb}\}$ the corner index in the checkerboard. The reprojection error is defined by the difference between the estimated and measured image point.

$$\boldsymbol{e}_{i,j,k} = \tilde{\boldsymbol{u}}_{i,j,k} - \boldsymbol{u}_{i,j,k}. \tag{7.4}$$

The estimated image point $\tilde{\boldsymbol{u}}_{i,j,k}$ in the distorted image is obtained by the refractive forward projection f_{rf} of the estimated checkerboard corner point $\boldsymbol{p}_{i,k} \in \mathbb{R}^3$ and a subsequent mapping f_{dc} from undistorted/rectified to distorted image coordinates

$$\tilde{\boldsymbol{u}}_{i,j,k} = f_{dc}(f_{rf}(\boldsymbol{p}_{i,k}, j, \boldsymbol{\pi}_p)). \tag{7.5}$$

The checkerboard point $\boldsymbol{p}_{i,k}$ in stereo camera coordinates is determined by

$$\boldsymbol{p}_{i,k}{=}\boldsymbol{\pi}_{cb,i} \oplus {}^{cb}\boldsymbol{p}_k, \tag{7.6}$$

where ${}^{cb}\boldsymbol{p}_k$ denotes the known point coordinates in the checkerboard frame.

If, like in our case, the distance d between the port and the camera is small and the optimisation algorithm moves the port behind the camera, using only the plain reprojection error can lead to undefined system states. In this case where $d < d_{\min}$, we use the alternative cost function

$$\begin{aligned} \boldsymbol{e}_{i,j,k} &= (f_{dc}(f_{rf}(\boldsymbol{p}_{i,k}, j, \boldsymbol{\pi}'_p)) - \boldsymbol{u}_{i,j,k})(1 + \Psi(\boldsymbol{\pi}_p)), \\ \boldsymbol{\pi}'_p &= (d_{\min}, 0, 0)^T, \\ \Psi(\boldsymbol{\pi}_p) &= c_1(|d - d_{\min}| + |y| + |p|), \quad c_1 \gg 1, \end{aligned} \tag{7.7}$$

which forces the optimiser to move the port back to a camera parallel pose in front of the camera.

Another extension of the reprojection error function $\boldsymbol{e}_{i,j,k}$ is required if the projected point $(u', v')^T = f_{rf}(\boldsymbol{p}_{i,k}, j, \boldsymbol{\pi}_p)$ does not lie in the undistorted/rectified image, and as a consequence, the distortion function f_{dc} is not defined. In this case, we use the cost function

$$\boldsymbol{e}_{i,j,k} = \begin{pmatrix} \max(0, -u', u' - w) + c_2 \\ \max(0, -v', v' - h) + c_2 \end{pmatrix}, \tag{7.8}$$

which determines the maximum distance of an image point to the edges of the visible image and forces the image points back into the visible camera views. Here, w and h represent the image width and height and c_2 a constant, which is greater than the maximum residual error caused by the noise in corner detection.

Our chosen stopping criterion applies, once the RMS residual error

$$\epsilon_{\text{res}} = \sqrt{\frac{e(\boldsymbol{x}, \boldsymbol{z})}{4nn_c}} \tag{7.9}$$

does not show any larger change between the iteration steps

$$\frac{\epsilon_{\text{res},i-1} - \epsilon_{\text{res},i}}{\epsilon_{\text{res},i-1}} < 0.0001\,\%. \tag{7.10}$$

7.3.3. Calibration of the dominating red and blue wavelengths

In chapter 6, we demonstrated the significance of the wavelength of light in 2D imaging and 3D reconstruction with flat port underwater cameras. With our wide-angle underwater short-baseline stereo camera, the wavelength dependent refraction of light is clearly visible as chromatic aberration and using a wrong wavelength can lead to close range reconstruction errors of more than 1 cm (see chapter 6).

We reduce the errors in imaging and reconstruction by exploiting the band-pass filters of the colour filter array of the camera and use only one colour channel to reduce the width of the wavelength spectrum. Due to the highest resolution of 50 % in the Bayer colour filter array, we use the green colour channel for the calibration of the poses of the port and the checkerboard.

Calibrating the dominating wavelengths rather than the index of refraction of water, as done by Telem and Filin [98], also has the advantage that calibrating the underwater camera is only required once. Instead of time-consuming recalibrations in place due to changing water properties, a one-time calibration of the underwater housing under laboratory conditions is sufficient and likely to be more accurate. Using the model of the index of refraction of water as introduced in chapter 6, our approach only requires the on-site measurement of the salinity, temperature, and pressure of water, which can be realised in only a fraction of a second with appropriate low cost sensors.

However, our attempts to calibrate the dominating wavelengths together with the pose of the port and the poses of the checkerboard (see section 7.3.2) revealed correlations between the wavelength, the distance of the port and the distance to the calibration target. The correlations prevented the optimiser from converging against the correct wavelengths. Accepting this limitation, we use the dominating wavelength of the green colour channel from the sensors data sheet (see section 6.3) for the calibration of the port and checkerboard poses, and subsequently calibrate the dominating wavelengths of the red and blue colour channel using the previously estimated port and checkerboard poses together with the dominating green wavelength from the data sheet.

Similar to the calibration of the port and checkerboard poses, the dominating red and blue wavelength, λ_r and λ_b, are estimated by minimising the total reprojection error

$$\lambda_i^* = \arg\min_{\lambda_i} e(\lambda_i, \boldsymbol{z}_i),\ i \in \{r, b\}. \tag{7.11}$$

Here, the image points $\boldsymbol{z}_i$ of the corresponding colour channel and the corresponding indices of refraction for the refractive forward projection f_{rf} are used. As an initial

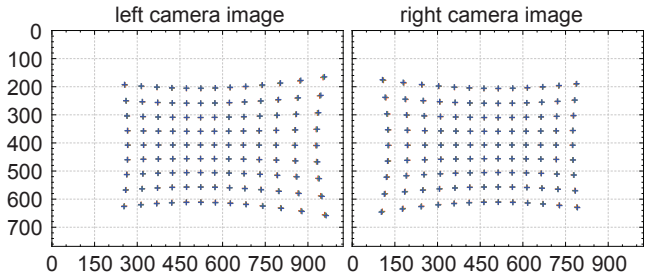

Figure 7.2.: The simulated stereo image(s) of the checkerboard from a fixed frontal view used to evaluate the calibration accuracy, execution time, convergence behaviour and necessary checkerboard pose distribution

guess, we use the dominating wavelength of the green colour channel to demonstrate that approximate initial guesses are also sufficient for convergence. If necessary, more accurate initial guesses, such as from the datasheet, can also be used.

7.4. Results

Our proposed accurate and efficient refractive forward projection method (see chapter 5) and the graph-based optimisation framework of Kümmerle et al. [117] with the Levenberg-Marquardt optimisation algorithm are deployed to compute and minimise the total reprojection error. For accurate numerical computation of the Jacobians, we exploit the high accuracy of our refractive forward projection method and choose a very small maximum 1-m-range back-projection error of 0.1 nm.

7.4.1. Simulation

To evaluate the convergence behaviour, calibration accuracy, calibration time and necessary checkerboard pose distribution, the proposed calibration is carried out with simulated data obtained from multiple runs with different numbers of measurements $n_z = \{1, 2, 5, 10, 25, 50, 100\}$. Although the proposed calibration method is designed for different checkerboard poses, we place the checkerboard at every measurement at the same almost front parallel pose at a close distance of 65 cm and a stereo camera relative orientation of $\theta_y = -93°$, $\theta_p = 1°$ and $\theta_r = -92°$ (yaw, pitch and roll). In the simulation, the barrel distortion of the ultra wide-angle lens is not modelled, that is $f_{dc} = 1$.

The resulting stereo image of the checkerboard with a dimension of $77 \times 55\,\mathrm{cm}^2$ is depicted in Fig. 7.2. To determine the statistical accuracy of the estimated port pose

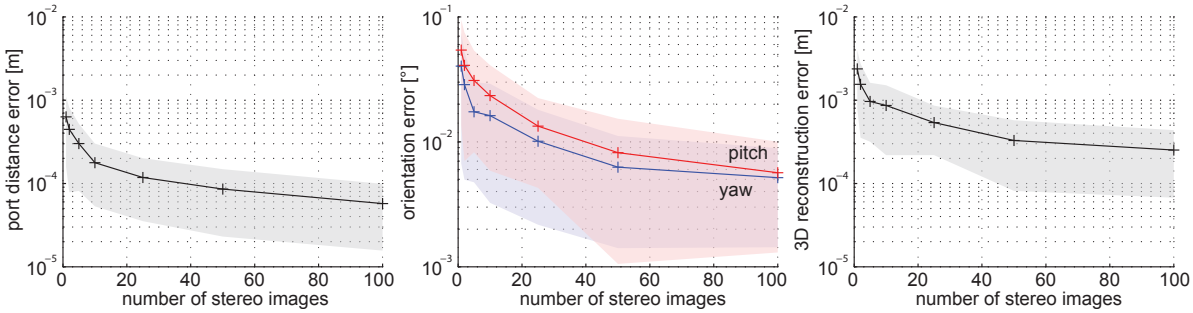

Figure 7.3.: Calibration with different numbers of stereo images of an almost front-parallel checkerboard: The mean errors with 1σ standard deviation of the calibrated port pose (distance, yaw and pitch angle) and of the resulting maximum 1-m-range 3D reconstruction error are shown. The corner detection accuracy was perturbed with bivariate Gaussian noise with a standard deviation of 0.5 px. The statistical results are based on 50 runs. Using 100 stereo images instead of 1, increases the accuracy of all measures by approximately one order of magnitude.

and the statistical execution time, every run with different n_z is executed 50 times. The error in corner detection is simulated by bivariate Gaussian noise with a standard deviation of 0.5 px.

The measured remaining errors in port orientation and distance, and the resulting maximum close range 3D reconstruction errors at a distance of 1 m are depicted as a function of the number of measurements n_z in Fig. 7.3. In all 350 runs, the optimiser converged successfully against the correct solutions. This indicates that the calibration in-water can be efficiently carried out from only one camera pose without the need to move the camera or the checkerboard, respectively. Thus, a small water tank could be used, for example.

Using $n_z = 100$ rather than $n_z = 1$ measurements increases the accuracy of all considered measures by approximately one order of magnitude. Our proposed calibration determines, with $n_z = 100$ stereo images, the port distance d with an average accuracy of 57 µm and the orientation in yaw and pitch direction, θ_y and θ_p, with an average accuracy of 5.2 m° or 5.7 m°, respectively.

If we compare these results ($n_z = 100$) with the single image monocular camera calibration of Yau et al. [72], our method estimates the port orientation ∼5.2 times more accurately and the port distance ∼3.5 times more accurately. For the comparison,

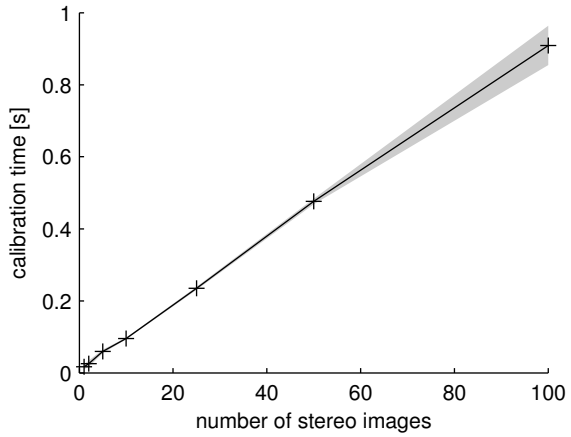

Figure 7.4.: Port pose calibration time: The mean and standard deviation are shown with respect to the number of stereo images. In contrast to prevailing literature, the graph shows the time efficiency of the proposed calibration method, which minimises the reprojection error. The execution time grows approximately linearly with the number of stereo calibration images and only requires 0.9 s for $n = 100$ image pairs.

we calculated the axis error by

$$e_{\boldsymbol{A}} = \arccos\left(\boldsymbol{A}^T \boldsymbol{R}_y(\theta_p)\boldsymbol{R}_z(\theta_y)\boldsymbol{A}\right)\frac{180^\circ}{\pi}, \tag{7.12}$$

where $\boldsymbol{A} = (1, 0, 0)^T$ denotes the axis of the axial camera [73] and $\boldsymbol{R}_y(\theta_p)$ and $\boldsymbol{R}_z(\theta_y)$ the rotation matrices for the rotation with yaw and pitch Euler angles.

The corresponding mean maximum error in reconstruction at a distance of 1 m amounts to 0.25 mm. In Fig. 7.4, also the processing time of our calibration is shown. The calibration time increases approximately linearly from 17 ms for $n_z = 1$ to only 0.9 s for $n_z = 100$ measurements.

As shown in Fig. 7.5, also the dominating red and blue wavelength, λ_r and λ_b, could be reliably and precisely estimated. Considering the calibration with $n_z = 100$ stereo images, λ_r could be estimated with a mean precision of 0.6 nm and λ_b with a mean precision of 0.2 nm.

7.4.2. Water tank experiment

In order to evaluate the proposed calibration method and the refractive underwater camera model under real conditions, the calibration method was tested in a water tank. We used a laminated printout of a calibration checkerboard with a similar size as in

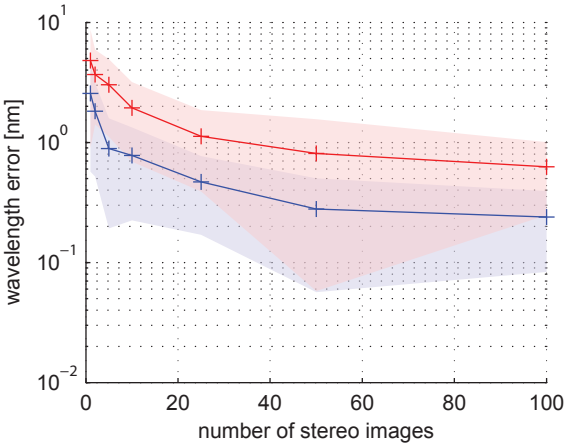

Figure 7.5.: Accuracy in the calibrated dominating wavelength of the red and blue colour channel: The mean error with the 1σ standard deviation is shown with respect to the number of stereo images.

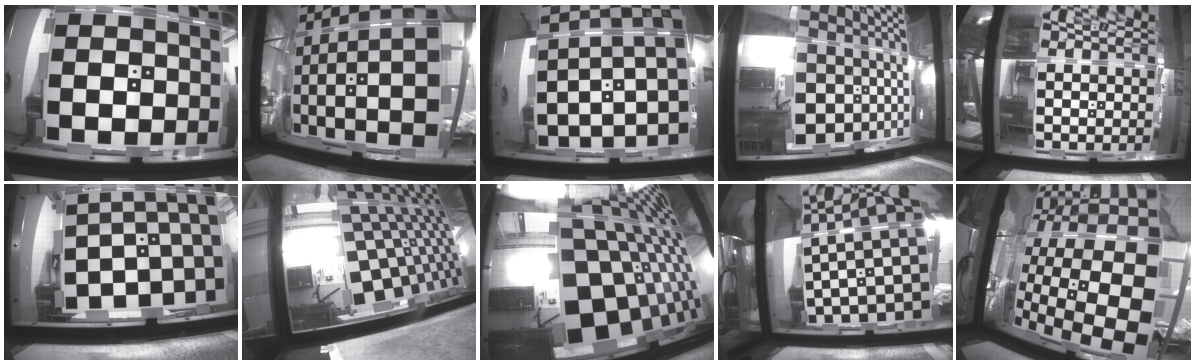

Figure 7.6.: Camera calibration in a water tank with multiple different camera views: The figure displays some example images of the submerged calibration checkerboard captured by the left camera.

the simulation, which was attached to a wall of the water tank. The wall of the water tank was assumed to be flat. To evaluate an appropriate camera view distribution under real conditions, two image sets with 80 stereo images each were captured from a fixed frontal camera view and from multiple different camera views. Some example images of the multi-view image set are depicted in Fig. 7.6. Before the checkerboard corners were detected in the green colour channel by standard corner detection [118], the green colour channel was extracted from the raw image according to the Bayer colour filter array, and the missing pixels were reconstructed with bilinear demosaicing. For the calibration, we measured a temperature of 18 °C, a salinity of 0.1 ‰ and an approximate water depth of 0.5 m. The camera housing was sealed at an approximate altitude of 13 m above sea-

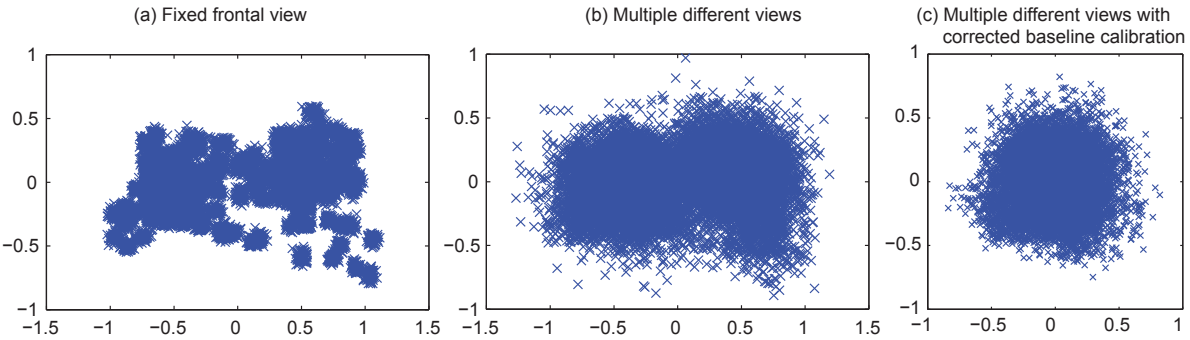

Figure 7.7.: Reprojection errors for different setups: (a) Multiple calibration images taken from a fixed frontal camera view ($\epsilon_{\text{res}} = 0.43\,\text{px}$); (b) Multiple camera images taken from uniformly distributed camera views ($\epsilon_{\text{res}} = 0.37\,\text{px}$); (c) Like (b) but with corrected baseline calibration ($\epsilon_{\text{res}} = 0.19\,\text{px}$). For each setup, 80 stereo calibration images were taken resulting in a total number of 18720 corners.

level. The port had a thickness of 12 mm. According to the data sheet of the camera sensor [53], the dominating green wavelength is $\lambda_g = 538\,\text{nm}$ (see section 6.3).

The resulting reprojection errors after calibration are shown in Fig. 7.7a and Fig. 7.7b for the single and multi-view setup. The fixed frontal view setup exhibits a slightly higher reprojection error $\epsilon_{\text{res}} = 0.43\,\text{px}$ than the multi-view setup with $\epsilon_{\text{res}} = 0.37\,\text{px}$. Instead of the expected normal distribution of the reprojection errors, the reprojection errors in Fig. 7.7a and Fig. 7.7b consist of two main point clusters. The point distribution in Fig. 7.7a deviates from a normal distribution, as the points are also grouped in many small distinguishable point clusters each resulting from the projections of a single checkerboard corner. The deviations from the expected normal distribution in Fig. 7.7a and Fig. 7.7b indicate inaccuracies in the models or its parameters. The small distinct point clusters in Fig. 7.7a are not attributable to plain measurement noise and indicate inaccuracies in the corner detection method used.

7.4.2.1. Proposed additional calibration of the stereo baseline

After an investigation of all model parameters used, we discovered that the RMS residual error can be almost halved to $\epsilon_{\text{res}} = 0.197\,\text{px}$ by decreasing the stereo camera baseline by $\sim$1 % from 12.07 cm to 11.95 cm ($-1.22\,\text{mm}$). Hence, we extended the port pose calibration step and included the baseline b as another calibration parameter

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} e(\boldsymbol{x}, \boldsymbol{z}), \tag{7.13}$$

where $\boldsymbol{x} = (\boldsymbol{\pi}_p, b, \boldsymbol{\pi}_{cb,1}, \ldots, \boldsymbol{\pi}_{cb,n})$. The resulting reprojection errors after an adaptation of the baseline are shown in Fig. 7.7c. The reprojection errors are now grouped in a single point cluster exhibiting a Gaussian-like bivariate distribution.

We analysed, why a modification of the baseline calibrated by the camera's manufacturer was necessary. So, we recalibrate the camera's baseline in air and left the remaining camera parameters unchanged. The calibration revealed a necessary baseline change of about $-0.5\,\text{mm}$. It is possible that the calibrated baseline differs from the original baseline of the manufacturer due to a specification error. The dimension of the checkerboard was accurately measured. Alternatively, perhaps the pinhole camera model, as part of the refractive underwater camera model, does not precisely mirror the actual refractive geometry and causes the errors observed.

7.4.2.2. Wavelength calibration

Based on the proposed port pose and baseline calibration and using the dominating wavelength $\lambda_g = 538\,\text{nm}$ of the green colour channel from the datasheet, the wavelength calibration estimated a value of $\lambda_r = 562\,\text{nm}$ for the dominating red wavelength and a value of $\lambda_b = 456\,\text{nm}$ for the dominating blue wavelength. In comparison with the calculated wavelength of the datasheet $\lambda_r' = 630\,\text{nm}$ and $\lambda_b' = 456\,\text{nm}$, λ_b could be estimated with high precision, whereas λ_r is almost 70 nm shorter than λ_r'. We suggest that the effect results from the attenuation of light in water being stronger for longer wavelengths. Also the light spectrum of the light sources, the filter characteristics of the infrared cut-off filter and the filter characteristic of the camera lens potentially contribute to the smaller value of λ_r. The very accurate estimate of λ_b indicates that λ_g from the data sheet represents a good initial guess, and that λ_g and λ_b were not significantly affected by the light attenuation.

7.5. Conclusion

In this chapter, a multi-view calibration framework for flat port underwater cameras was proposed. This framework is similar to standard in air multi-view camera calibration techniques with a flat checkerboard calibration target. The proposed method requires a pre-calibrated in-air stereo camera and estimates the orientation and distance of the port and the dominating wavelength of the red and blue colour channel. The method exhibits a high accuracy due to the amount of calibration data, which can significantly exceed the amount of data of prior underwater camera calibration techniques. Compared with

the recently published single image monocular camera calibration method of Yau et al. [72], which requires a special LED calibration target, our method achieves a ∼5.2 times higher accuracy in estimating the port orientation and a ∼3.5 times higher accuracy in estimating the port distance if 100 simulated stereo calibration images are used. We showed that in spite of hundreds of calibration images, accurate and efficient underwater camera calibration with reprojection error minimisation is possible in only a few seconds. This is significantly faster than the 5 minutes or several hours required for the calibration with one or multiple image pairs, respectively, required in the calibration method of Sedlazeck and Koch [75]. Practical water tank experiments revealed the limitations of the standard physics-based model and the limitations of standard corner detection for underwater camera calibration. To compensate for these limitations, we recommend the recalibration of the stereo baseline and the calibration with different camera views. In this way, we almost halved the reprojection error to 0.19 px.

8. Underwater 3D reconstruction accuracy

8.1. Introduction

In this chapter the overall reconstruction accuracy of our wide-angle flat port underwater short baseline stereo camera from a single camera view is evaluated. The chapter shows how well the physics-based underwater camera model with the proposed integration of the index of refraction models (see chapter 6) and the proposed underwater camera calibration method with additional baseline recalibration (see chapter 7) performs under real-world conditions. This evaluation is based on a propose method to evaluate the 3D distortion of the reconstructed object space.

8.2. Related literature

There is no prior research on the reconstruction accuracy of wide-angle flat port underwater short-baseline stereo cameras under real-world conditions with varying object distances.

In 1986, Fryer and Fraser used the perspective camera model in combination with the standard lens distortion model to imitate the effects of refraction. In a bundle adjustment approach, the authors calibrated the principle point, the focal length and the pose of the camera and the pose of a 3D calibration target. The calibration frame had about 25 target points. Fryer and Fraser evaluated the reconstruction accuracies of both an analogue in-air camera embedded in a flat port underwater housing (camera 1) and an analogue wide-angle flat port underwater camera (camera 2) specifically designed for the use in-water. This wide-angle underwater camera had a HFOV of 66°, which is comparable to the 70° HFOV of our underwater camera. For their calibration, four images from different poses but with a common object distance of 2.5 m or 1.9 m, respectively, were taken resulting in about 90 to 100 image points. The reconstruction accuracy was

determined from the covariance matrices of the estimated target points. For camera 1, Fryer and Fraser obtained a reconstruction accuracy with a mean standard deviation of 0.18 mm in lateral direction and 0.26 mm in depth direction, and for camera 2 0.23 mm in lateral direction and 0.4 mm in depth direction.

The results of Fryer and Fraser are very accurate and comparable to our results. However, there are significant differences to our approach. Fryer and Fraser used the results of the calibration to evaluate the reconstruction accuracy. But using the calibration results for evaluating the reconstruction accuracy is not recommended, as the reconstruction may be only accurate for the calibrated object points alone. So we used a completely different image set with also significantly more image points. Moreover, we reconstructed the object points from only a single view. A disadvantage of Fryer and Fraser's camera model is the limitation to constant object distances, which does not allow the reconstruction of scenes with significant variation in depth.

In 1997, Li et al. [95] evaluated the reconstruction accuracy of a stereo camera with two separate flat port underwater housings and a baseline of 40 cm. The camera was placed 2-3 m in front of the a three-dimensional frame, which had a dimension of $1.4 \times 1.4 \times 0.7\,\mathrm{m}^3$ and was equipped with control points from which 15 were used and placed with a precision of 1 mm in the frame. The pose of the stereo camera was determined in an empty water tank using the measurement data of the camera itself. The evaluation of the reconstruction accuracy was based on a single stereo image with a resolution of 768×494 px and resulted in a root mean square reconstruction error of 6 mm in lateral and of 7 mm in depth direction.

Even though Li et al.'s and our stereo camera and experimental setup are not directly comparable, the reconstruction accuracy is significantly higher for of our wide-angle underwater stereo camera with a shorter baseline and a higher resolution. For our close range experiment of about 65-105 cm, we measured a maximum mean error of only 0.2 mm in baseline direction, 0.5 mm in vertical direction and 1.8 mm in depth direction. Since Li et al. used the same refractive camera model, we attribute the higher accuracy of our system to our proposed camera calibration method (see chapter 7). Our evaluation is also more precise, because we considered a 3D grid of about 1050 points, where each point was reconstructed 10 times. The 3D grid also enabled a straight forward evaluation of the distortion in the reconstructed 3D space.

The stereo-rig of Shortis and Harvey [101] from 1998 to measure the size of fish is the only underwater stereo camera in literature whose reconstruction accuracy was actually evaluated under real-world conditions. It consisted of two camcorders with separate

underwater housings and a large baseline of 1.4 metres. Although the camcorders had digital sensors, an analogue recording of the images was made on tape and subsequently digitized with a frame grabber. The perspective camera model in combination with the standard lens distortion model was used to model the effects of refraction. In the reconstruction experiment in a pool, 16 fish silhouettes with a size of 10-49 cm were used, and measured from multiple different camera distances. From each camera view, multiple measurements were taken and averaged. Shortis and Harvey noticed a clear degradation in reconstruction accuracy for large angles, but also a systematic error in underestimating the fish length. The authors obtained a mean reconstruction error of 1.3 cm.

The reconstruction accuracy in the order of centimetres of Shortis and Harvey's long-baseline stereo camera is relatively large compared to our short-baseline stereo camera with reconstruction accuracies of millimetres to sub-millimetres. We would attribute the large reconstruction errors to the use of an inaccurate perspective camera-based underwater camera model, the loss in image quality due to the conversions between digital and analogue image data, and the manual selection of image points.

In 2008 and 2012, Treibitz et al. [68, 71] evaluated the reconstruction accuracy of a close-range size measurement system for ranges of about 0.5 m to 1.5 m, which consisted of a monocular underwater camera with in-situ adjustable zoom and focus. A physics-based refractive camera model was used, but the thickness and orientation of the port were simplified by single interface approximation and parallelism to the camera's image plane. The index of refraction of water was also simplified using a value of $n_w = 4/3$. The reconstruction accuracy was evaluated with objects of known length that were placed in the front parallel plane, which the camera was calibrated for. The evaluation revealed a size measurement error of up to 1.2 cm.

The considered object distances in Treibitz et al.'s close-range size measurement system are similar to our evaluated ranges. Since Treibitz et al. use a monocular underwater camera, a direct comparison with our underwater stereo camera is difficult. However, the large errors in size measurement of up to 1.2 cm are relatively large compared with our maximum mean lateral reconstruction error of less than 0.5 mm. We attribute the large reconstruction errors primarily to the simplification of parallelism between the port and the camera.

In 2010, Telem and Filin [98] evaluated the accuracy of the underwater photogrammetry system based on a single narrow field of view flat port underwater camera with a physics-based refractive camera model and approximated refractive forward projection.

For the reconstruction, structure from motion in combination with reference information from marked planar plates and objects of known dimension was used. Telem and Filin evaluated the photogrammetric system in a pool and the open sea and achieved high reconstruction accuracies. In the pool experiment with an object distance of about 4-5 m, the RMS error in measuring the length of markers was ∼2 mm, and in the open sea 0.3 mm with object distance of 2-3 m. Furthermore, the deviation of reconstructed points of objects with known geometry was determined. In the swimming pool experiment, the measured mean deviation from a straight line was 0.7 mm and from a plane 0.8 mm, with slightly better results in the close-range open sea experiment.

The reconstruction accuracy of Telem and Filin's photogrammetric system with a monocular camera is comparable to our results or slightly better. In contrast to Telem and Filin, however, we evaluate the reconstruction accuracy only from a single (stereo) camera view and do not use additional reference objects placed in the scene. The high reconstruction accuracy of Telem and Filin's system can be attributed to the large baselines and multiple different camera views. Whereas Telem and Filin mainly focused on the capability of their photogrammetric system to measure distances or to reconstruct geometric primitives (lines, planes), our experimental setup also allows the evaluation of the distortion of the reconstructed 3D space.

In 2012, Kang et al. [81] reconstructed objects in a water tank from multiple in-air views. The multiple views of the objects were obtained either with a single camera and a turntable placed underneath the object, or by 8 in-air cameras symmetrically placed on a 180° circular arc around a water tank with multiple flat walls. Kang et al. avoided the physics-based refractive camera model, as it is allegedly too time-consuming. Instead, the reconstruction accuracy was evaluated based on the perspective camera model with an adjusted focal length and adjusted distortion coefficients. The camera(s) had a resolution of 1032×776 px and a narrow field of view of 32°. The reconstructed objects had a dimension of $\sim 20\,\mathrm{cm}^3$ and were placed in a range between 40 cm to 75 cm in front of the camera(s). The ground truth of the object structures was determined in air on a turntable and later aligned with the corresponding reconstructions. For the reconstruction of the scene, bundle adjustment and scale-invariant feature transform (SIFT) [119] were used. By this means, Kang et al. achieved a high reconstruction accuracy between 0.39 mm to 0.69 mm.

Kang et al. achieved very high reconstruction accuracies, which are comparable to our results. This was even achieved with a perspective camera-based model. However, the setup used does not enable a general statement on the reconstruction accuracy of

perspective camera-based underwater camera models, as the object distance did not significantly change due to the experimental setup, which is however a fundamental property of flat interface underwater vision. The distances of the object points were only slightly changed due to the rotations of the object or the views, respectively. We attribute the high accuracy of Kang et al.'s work mainly to the high number of views from very different camera poses (located on a circular loop). In contrast to that, we evaluated the reconstruction accuracy from a single stereo camera view and for different object distances.

8.3. Proposed evaluation of the underwater 3D reconstruction accuracy

8.3.1. Experimental setup

To evaluate the 3D reconstruction accuracy of our underwater stereo camera, we attached a checkerboard to the inner wall of a water tank (see Fig. 8.1). A carrier on top of the water tank held the camera and constrained its motion to a linear translatio n $s \cdot \boldsymbol{t}$, with $s \in \mathbb{R}$, $\boldsymbol{t} \in \mathbb{R}^3$, which was approximately perpendicular to the checkerboard. To evaluate the reconstruction accuracy at different distances, several images at fixed distances between $\sim$65 cm and $\sim$105 cm with a raster of $t = |\boldsymbol{t}| = 5$ cm were taken. The displacement t of the carrier was determined with an accuracy of $\sim$0.5 mm. Some examples images of the resulting $n_d = 9$ different camera views are displayed in Fig. 8.2.

8.3.2. Constrained calibration of the checkerboard poses

The experimental setup constrains the orientation and distance between camera poses. However, the poses $\boldsymbol{\pi}_{cb,1}, \ldots, \boldsymbol{\pi}_{cb,n}$ of the checkerboard relative to the camera and the direction of the displacement vector $\boldsymbol{t} \in \mathbb{R}^3$ are still unknown.

In contrast to Li et al. [95], we did not determine the ground truth poses of the checkerboard in an empty water tank using the reconstruction capabilities of the in-air stereo camera. We made this decision for several reasons. Firstly, opening and closing the housing would have made a recalibration necessary. Secondly, due to the mounting mechanism of the housing, it is not easy to open it without changing its tilt. Thirdly, the long aluminium camera holder attached to the carrier was not solid enough to keep

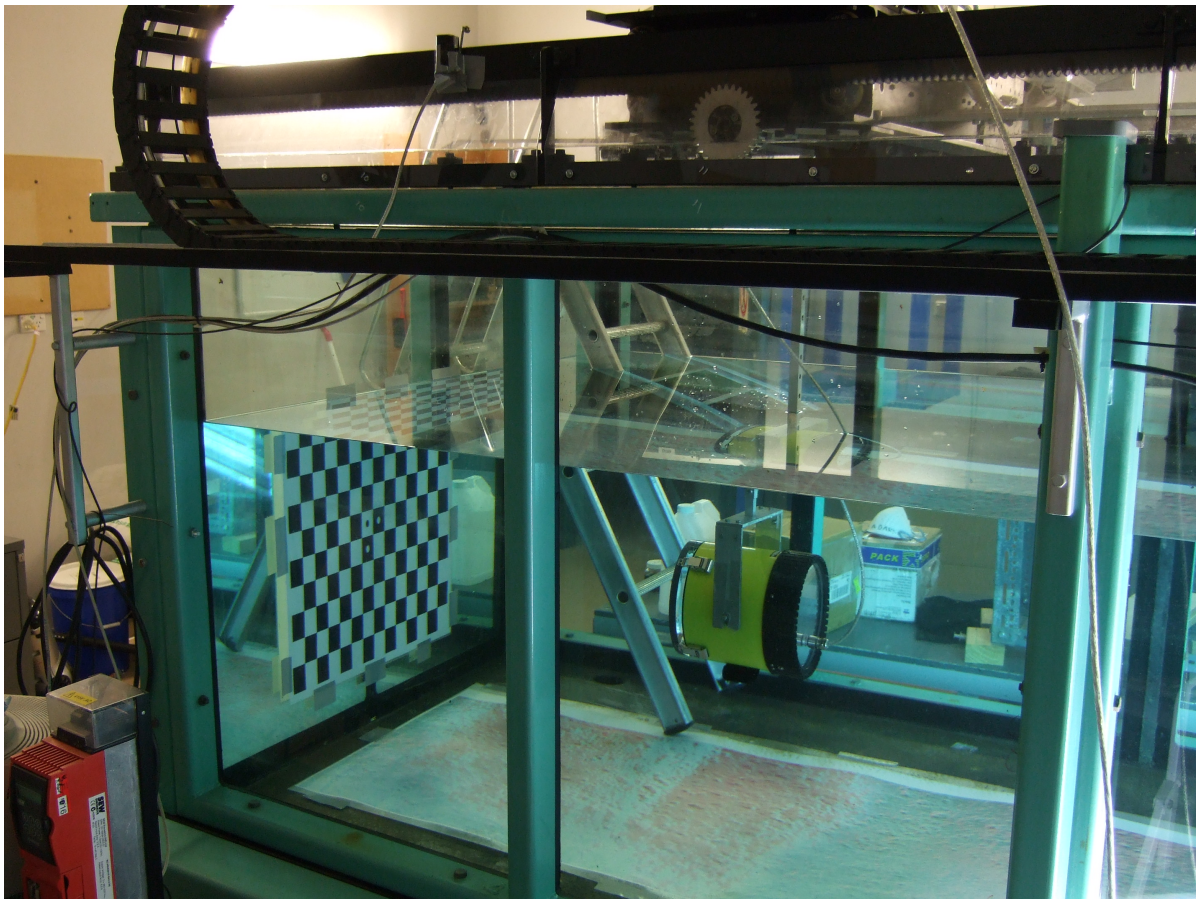

Figure 8.1.: The experimental setup to determine the reconstruction accuracy of the underwater camera: The figure shows the checkerboard pattern, which was attached to a wall of the water tank and the underwater camera, which was held by a movable carrier on top of the water tank.

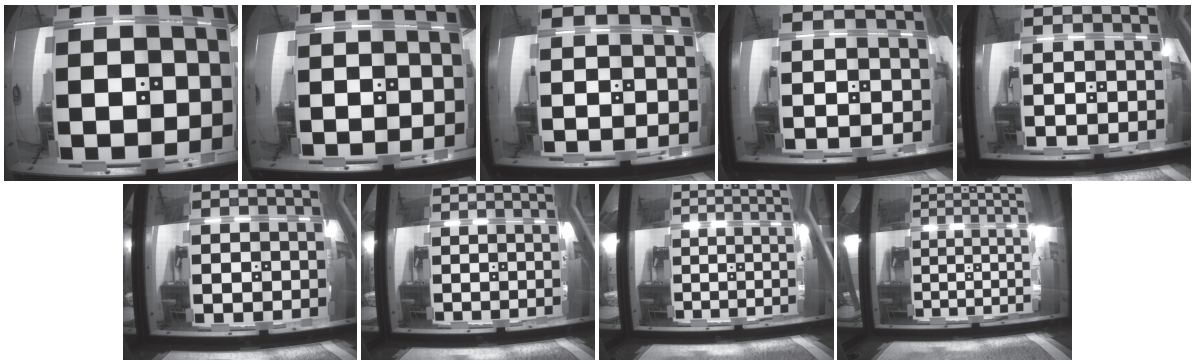

Figure 8.2.: Evaluation of the 3D reconstruction accuracy with a checkerboard at different distances: The left images of the stereo image pairs are shown.

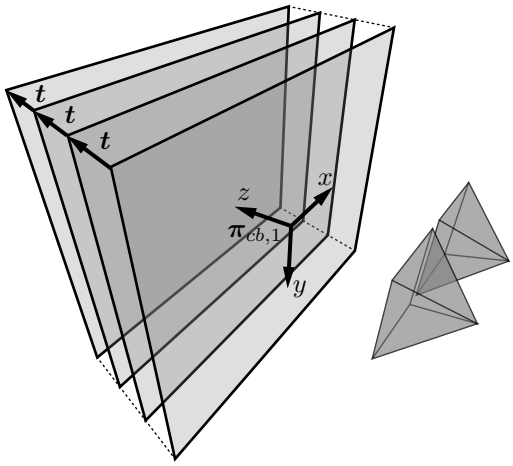

Figure 8.3.: Constrained calibration of the checkerboard poses: Since the ground truth of the checkerboard with respect to the camera is unknown, multiple images of the checkerboard are taken from different distance to create a constrained reference grid of points.

the housing in a very precise position with and without water. Only small changes in the camera pose due to changing buoyancy would have significantly diminished the accuracy of the experimental setup. For that reason, we decided to keep the underwater camera closed and at a constant water depth.

To determine the pose of the checkerboards, we exploit the reconstruction capability of the underwater stereo camera itself, but additionally use the motion constraints described above. Based on a camera centric coordinate system, the approach consists of two calibration steps in which firstly the pose of the closest checkerboard $\boldsymbol{\pi}_{cb,1}$, and secondly the translation vector $\boldsymbol{t}$ between the checkerboards is calculated (see Fig. 8.3). Even though it is not possible with our experimental setup to determine the reconstruction accuracy with respect to the stereo camera, our method allows the straight forward evaluation of the relative accuracy in 3D object space.

For the estimation of the closest checkerboard pose $\boldsymbol{\pi}_{cb,1}$, we use $n_{z,1} = 100$ stereo image pairs and minimise the total reprojection error of the detected checkerboard corners

$$\boldsymbol{\pi}^*_{cb,1} = \underset{\boldsymbol{\pi}_{cb,1}}{\arg\min}\, e(\boldsymbol{\pi}_{cb,1}, \boldsymbol{z}_1). \tag{8.1}$$

Similar to the port pose calibration of the underwater camera (see chapter 7), the checkerboard corners $\boldsymbol{z}_1$ are detected in the distorted, demosaiced green colour channel of each image. The green colour channel is preferred to the red and blue colour

channel due to its higher resolution in the Bayer colour filter array pattern.

In the second calibration step the direction of the translation vector $\boldsymbol{t}$ between the checkerboard poses is calculated. With respect to the orientation of the checkerboard pose $\boldsymbol{\pi}_{cb,1}$, $\boldsymbol{t}$ is parameterised by the pitch and roll Euler angles, $\boldsymbol{x}_t = (\theta_p, \theta_r)^T$. With the given translation distance t and the rotation matrix $\boldsymbol{R}(\boldsymbol{x}_t)$, we have

$$\boldsymbol{t}=\boldsymbol{R}(\boldsymbol{x}_t) \begin{pmatrix} 0 \\ 0 \\ t \end{pmatrix}. \tag{8.2}$$

Also here, θ_p and θ_r are estimated by minimising the total reprojection error

$$\boldsymbol{x}_t^* = \arg\min_{\boldsymbol{x}_t} e(\boldsymbol{x}_t, \boldsymbol{z}_2). \tag{8.3}$$

In this step, we use $n_{z2} = 10$ stereo images of each checkerboard. The checkerboard corners $\boldsymbol{z}_2$ are again extracted from the green, demosaiced colour channel. The 3D position of the checkerboard corners are defined by

$$\boldsymbol{p}_{k,l}=\boldsymbol{\pi}^*_{cb,1} \oplus ({}^{cb}\boldsymbol{p}_k + (l-1)\boldsymbol{t}), \tag{8.4}$$

where k denotes the index of the corner in the checkerboard, l the index of the checkerboard and ${}^{cb}\boldsymbol{p}_k$ the coordinates of a checkerboard corner in the checkerboard frame.

8.3.3. Comparing the reconstructed 3D object points with the 3D point grid

The constraint calibration of $\boldsymbol{\pi}_{cb,1}$ and $\boldsymbol{t}$ in section 8.3.2 results in a regular grid of 3D checkerboard points. The grid is used as reference for the evaluation of the reconstruction accuracy and shows in a straight forward way how well the refractive camera model can compensate for refractive 3D distortion.

For the evaluation of the reconstruction accuracy, we consider reconstruction from single measurements. For that purpose, we deploy the proposed reconstruction method from section 6.2, which determines the positions of the object points by minimising the reprojection errors.

8.4. Results

The reconstruction results of our wide-angle flat port underwater stereo camera (see Table 2.1) are visualised in Fig. 8.4. The figure shows the reference grid with $n_g = 1053$ green spheres and the reconstructed points with $n_g n_{z2} = 10530$ red spheres from three different orthographic views. Each sphere has a diameter of 5 mm. The pose $\boldsymbol{\pi}_{cb,1}$ of the grid or of the closest checkerboard, respectively, is depicted. Here, the x-axis and y-axis denote the direction of the columns and the direction of the rows of the checkerboard and the z-axis its normal directed away from the camera. The checkerboard had a dimension of $77 \times 55\,\mathrm{cm}^2$.

In all views, the reconstructed points are mostly very close to the reference points of the grid. In general, both point groups overlay each other indicating low noise but also a precise reconstruction of the scale, which was achieve by the proposed recalibration of the baseline as described in chapter 7.

The front view does not show noticeable errors in reconstruction. The underwater camera accurately reconstructed the grid points in lateral direction to the optical axis.

Also the top view shows that the distances for most of the points could be accurately reconstructed. The distance in z-direction from the vertical lines in y-direction of the point grid is kept small for most of the reconstructed points. A distortion in depth can be observed for the outer vertical lines (e.g. 1, 12 and 13; see Fig. 8.4) of the checkerboard, which have a larger angle to the optical axes of the cameras. Here, the close points were reconstructed at a too short object distance and the farther points at a too large distance to the camera. Moreover, for most of the reconstructed points, the small deviations from the vertical lines kept approximately constant with distance and the vertical curves defined by these points did not change their shape significantly. This is particularly visible for the vertical lines 6, 7 and 13 (see Fig. 8.4). It indicates that the underwater reconstruction even detected the non-planar surface of the checkerboard or of the wall to which it was attached. As shown in Fig. 8.2, the thin laminated printout of a checkerboard pattern was only attached at the edges to the water tank using adhesive tape, and so bending of the checkerboard could not be completed avoided.

The orthogonal view of the grid from the right side of the camera confirms the capability of the underwater camera to reveal deviations of the checkerboard from an exactly planar shape. It also confirms the slight distortion in depth at large angles in the direction of the baseline. The shape of each reconstructed horizontal grid line changes with distance from a slight concave to a slight convex shape.

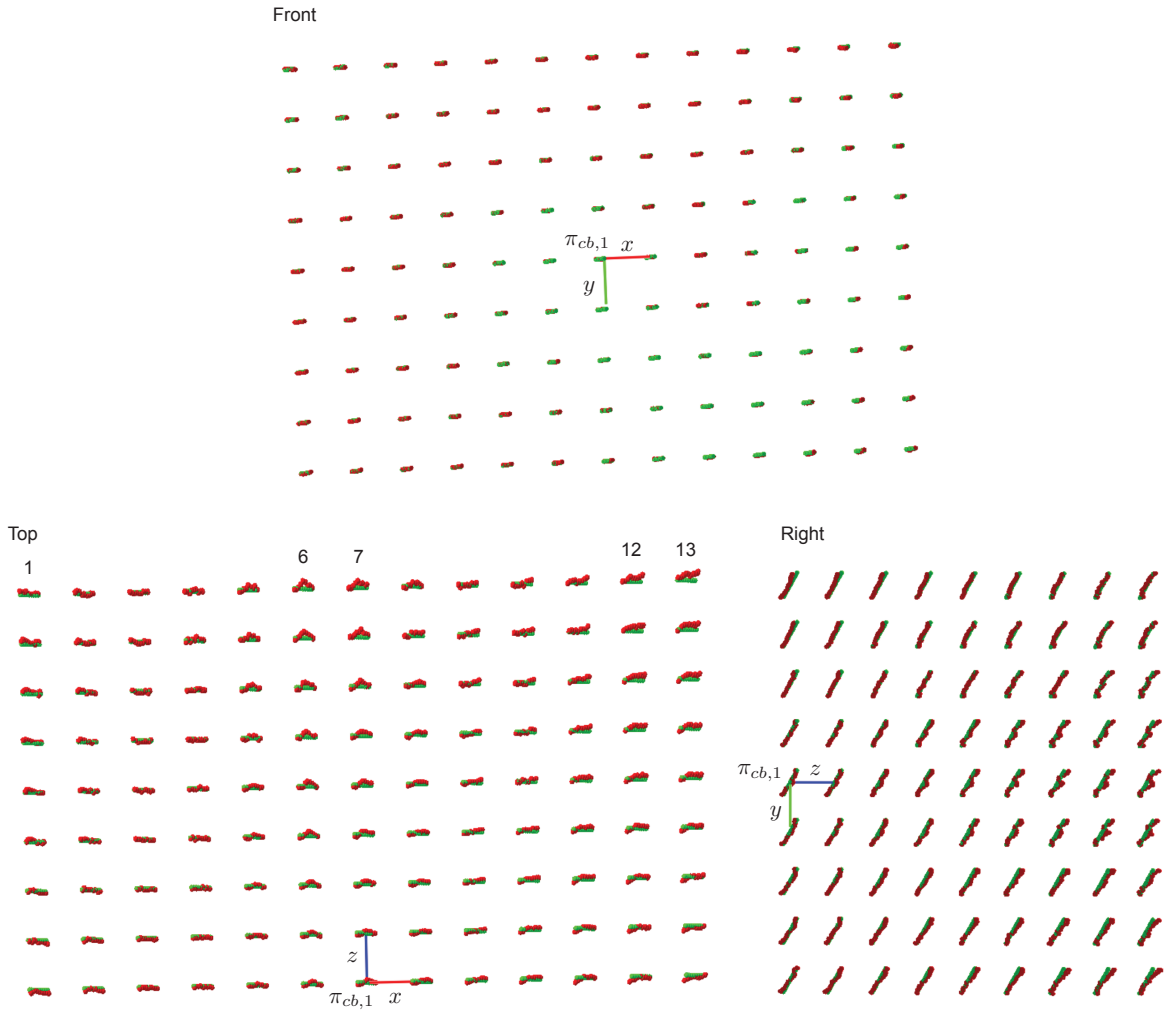

Figure 8.4.: Visual evaluation of the underwater reconstruction accuracy using a reference grid: The image shows three orthographic views of the reference grid (green spheres) and of its reconstruction (red spheres). The spheres have a diameter of 5 mm.

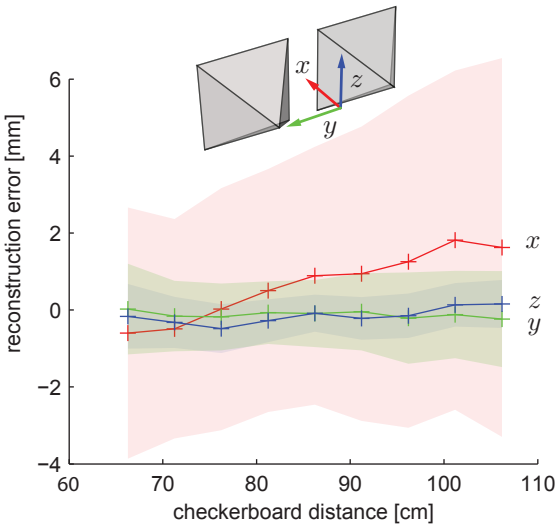

Figure 8.5.: Mean reconstruction errors with 2σ bounds

For a quantitative description of the results, the mean errors in x-, y- and z-direction of the stereo camera frame are depicted in Fig. 8.5 for different checkerboard distances together with the 2σ standard deviation. The frame of the stereo camera is also depicted in Fig. 8.5. The x-axis of the stereo camera frame is parallel to the optical axes of the camera, the y-axis is parallel to the baseline and the z-axis denotes the vertical direction. The mean reconstruction error in y-direction was smallest and varied between $-0.23\,\text{mm}$ and $0.02\,\text{mm}$, followed by the error in z-direction with values between $-0.48\,\text{mm}$ and $0.16\,\text{mm}$. The errors in both directions did not significantly correlate with the checkerboard distance. As expected, the mean errors in x-direction are largest and varied between $-0.6\,\text{mm}$ and $1.8\,\text{mm}$. The error increased in x-direction with growing distance. The 2σ standard deviation in z-direction kept almost constant with distance and ranged between $0.48\,\text{mm}$ and $0.84\,\text{mm}$, followed by the standard deviation in y-direction with $0.81\,\text{mm}$ to $1.2\,\text{mm}$. The standard deviation in y-direction increased slightly with distance. As expected, the largest standard deviation and the largest increase with distance was measured for the x-direction. Here, the 2σ standard deviation ranged between $2.8\,\text{mm}$ and $4.9\,\text{mm}$.

With respect to the long-term goal of the project, in which small marine organisms on the hull will be reconstructed, it is sufficient to consider only the mean reconstruction errors of the closest checkerboard. In this case, the closest checkerboard would represent the hull and the reconstruction errors are measured with respect to this hull. For the reconstructions of the objects points relative to the hull or the checkerboard, re-

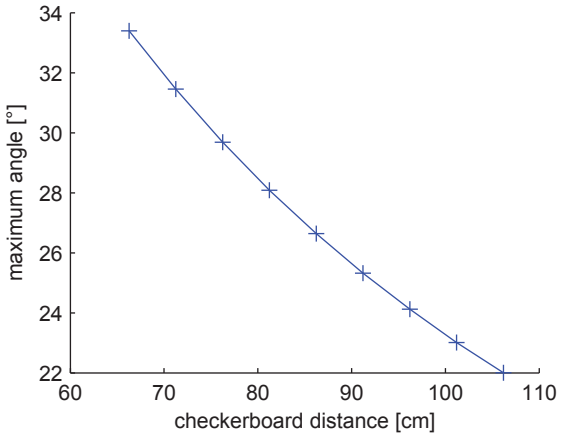

Figure 8.6.: The approximate field of view covered by the checkerboard experiment: The figure shows the maximum angles to the x-axis of the stereo camera for different checkerboard distances.

spectively, we obtained a mean maximum reconstruction error of 0.6 mm in x-direction, 0.22 mm in y-direction, and 0.16 mm in z-direction over the whole checkerboard. Since a single organisms covers a much small area, these errors will significantly decrease for the reconstruction of marine species.

The method chosen enabled the evaluation of the reconstruction accuracy of the underwater camera for a narrow viewing volume in front of the camera. Fig. 8.6 shows the maximum angles of the reconstructed points to the x-axis of the stereo camera frame for different checkerboard distances.

8.5. Conclusion

A novel reconstruction accuracy of a wide-angle flat port underwater short-baseline stereo camera was evaluated under real-world conditions. A method for straight forward evaluation of the distortion of the reconstructed 3D space was proposed. It is based on a 3D grid of points, which is generated with a checkerboard. The reconstruction accuracy for short ranges between 65-105 cm where the distance dependency of refractive 3D distortion is high, was evaluated. The experiments showed that the mean reconstruction accuracy in lateral direction does not significantly change with distance and is better than 0.5 mm. The mean reconstruction error in depth with respect to the closest grid layer changed with distance from -0.6 mm to 1.8 mm. The proposed evaluation method also revealed small distortions in the reconstructed object space. The reconstruction

accuracy of our underwater stereo camera is comparable with the structure from motion based underwater reconstruction methods of Fryer and Fraser [64], Telem and Filin [98] and Kang et al. [81]. However, we reconstructed the scene only from a single camera pose and neither require any additional reference targets [64, 98] in the scene nor restricted the camera motion to certain object distances [64, 81].

With respect to the long term goal of the project, the results are very promising. If we consider the reconstruction accuracy of the hull of a ship at a distance of about 65 cm, the proposed system reconstructs points on the hull with an accuracy of 0.6 mm in depth direction, 0.22 mm in baseline direction, and 0.16 mm in vertical direction of the camera. These accuracies apply to a relative large ship hull surface area of $77 \times 55\,\mathrm{cm}^2$. Since the dimensions of marine organism on the hull are much smaller, the accuracies for the reconstruction of marine species would be still much higher.

9. Improving the single interface approximation

9.1. Introduction

The thickness of the flat port of an underwater housing, the thickness of the wall of a water tank or, in general, the thickness of a flat glass plate has been incorporated into the physic-based refractive camera model in two different ways. Research groups are split into two camps, where on the one hand, [72, 71, 68, 98] approximate the glass plate by a single interface and, on the other hand, [78, 79, 77, 80, 100, 75, 2, 73, 59, 95, 1] explicitly model the thickness of the glass plate by two parallel interfaces.

The usage of the single interface approximation seems appealing, as the complexity to compute refractive forward projection reduces to a simplified quartic equation, which can be solved in an accurate closed-form [83]. As we have shown in chapter 5, though, the computation of the quartic equation is less efficient than solving refractive forward projection by a back projection-based or lookup table-based approach. Hence, efficiency is no longer a valid reason for choosing the single interface approximation. In this section, we investigate the accuracy in projection and reconstruction of the single interface approximation. In this context, a more accurate calibration technique is proposed, which incorporates the viewing volume of the stereo camera and considers two separate calibrations of the single interface to minimise the total projection and reconstruction error.

9.2. Related literature

To date, almost no publication evaluated the accuracy of the single interface approximation. Treibitz et al. [68] mentioned that a small radial shift of the ray in water is induced by the thickness of the port. A small incident angle $\theta_w = 20°$ of the light ray in water yielded a shift of $\sim$0.28 mm and was ignored. The error for larger θ_w was not considered.

Telem and Filin [98] studied the effect of the thickness of the port on their model with an adapted focal length, and showed how the virtual interface needs to be arranged to precisely simulate the refraction of a single point. Agrawal et al. [73] demonstrated the approximation accuracy of single interface approximation. However, they averaged the indices of refraction of water and glass for the virtual single interface, which is inaccurate as we will show. Furthermore, they calibrated the pose of the single interface with a calibration target and evaluated the resulting reprojection error. However, this calibration method is only optimal with respect to the limited number of calibration images, and so does not consider a more comprehensive viewing volume.

9.3. Proposed method

In this section, we firstly show that the indices of refraction of the port and water should not be averaged. Secondly we demonstrate that a single interface model is only equivalent to a thick port model for object points having the same radial distance in the distorted image. In addition to that, we propose a single interface approximation for underwater stereo cameras, which achieves a higher accuracy in refractive 2D forward projection and 3D reconstruction by incorporating the stereo viewing volume. In this context, we also present visualisations of the refractive stereo viewing volume and visualisations of the common image regions as seen by both cameras. Additionally, we demonstrate that optimal accuracies in refractive 2D forward projection and 3D reconstruction can be only realised by two separate calibrations. Lastly, the possible 2D and 3D accuracies with respect to the port thickness are demonstrated.

9.3.1. Equivalent single interface model for a single radial distance

In this section, we explain why averaging of the indices of the port and water [73] is less accurate and why an equivalent single interface model only exists for a single radial distance, assuming the port is parallel to the camera image plane. Therefore, we use the optimal interface shift equation of Telem and Filin [98].

The single interface approximation is displayed in Fig. 9.1. The figure shows the propagation and refraction of a light ray through a thick port. Considering only the light rays outside the port, it is possible to precisely simulate the system with only a single interface so that the light rays in water and air are identical. According to Snell's

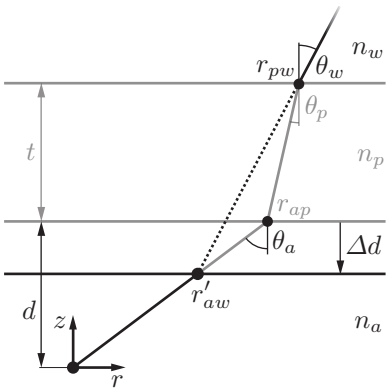

Figure 9.1.: Single interface approximation: A flat port can be precisely simulated by a single interface if only one light ray with the incident angle θ_w and the refracted angles, θ_p and θ_a, is considered. In this case, the single interface lies in between the air-port interface and the camera.

law

$$n_a \sin\theta_a = n_p \sin\theta_p = n_w \sin\theta_w, \tag{9.1}$$

where n_a, n_p and n_w denote the indices of refraction of air, the port and water, the angles θ_a and θ_w of the rays in air and water can be expressed independently from the index of refraction n_p of the port. Therefore, n_p should not be incorporated into the single interface approximation as done in [73]. If n_p was not ignored the virtual and correct ray in water would not be parallel and would cause an error growing with the scene distance.

In comparison to a single interface, a thick port causes a displacement of the ray in water. This displacement can be exactly imitated by a single interface placed on the intersection point r'_{aw} of the light rays in air and water (Fig. 9.1). The optimal position of the interface, which is parallel to the port, is always located between the port and the camera and can be mathematically described. The distance Δd between the virtual interface and the port can be derived by considering the radial coordinates r_{ap} and r_{pw} of the actual refraxes and the radial coordinate r'_{aw} denoting the intersection point of the ray in water with the single interface. From [98] follows:

$$\underbrace{\Delta d \tan\theta_a}_{r_{ap}-r'_{aw}} + \underbrace{t \tan\theta_p}_{r_{pw}-r_{ap}} = \underbrace{(\Delta d + t)\tan\theta_w}_{r_{pw}-r'_{aw}} \tag{9.2}$$

The equation can be expressed as a function of the refracted angle θ_a in air, the indices

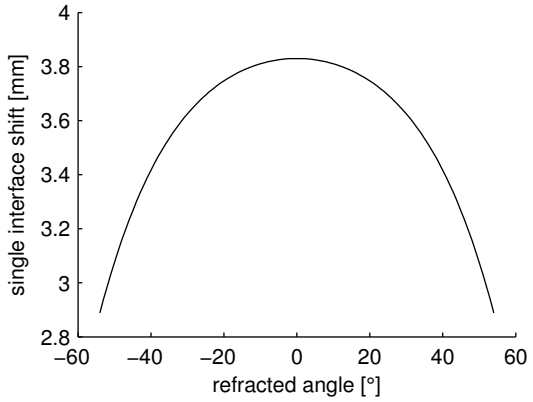

Figure 9.2.: Single interface approximation: The optimal interface shift Δd to a 12 mm thick port to precisely simulate the light rays in air and water.

of refraction n_a, n_p, and n_w and the port thickness t, where

$$\tan\theta_i = \frac{\sin\theta_i}{\cos\theta_i} = \frac{n_{ai}\sin\theta_a}{\sqrt{1-(n_{ai}\sin\theta_a)^2}}, \tag{9.3}$$

and $n_{ai} = n_a/n_i$, with $i \in \{p, w\}$. Solving (9.2) for Δd and removing the singularity at $\theta_a = 0$, we obtain

$$\Delta d = \begin{cases} t\frac{\tan\theta_w - \tan\theta_p}{\tan\theta_a - \tan\theta_w}, & \text{if } \theta_a \neq 0, \\ x \in \mathbb{R}, & \text{else} \end{cases} . \tag{9.4}$$

In Fig. 9.2, the optimal interface shift Δd is plotted as a function of θ_a using the refractive indices $n_a = 1.000273$, $n_p = 1.4943$ and $n_w = 1.3347$, and the port thickness $t = 12$ mm. The figure shows a U-shaped curve having its maximum at small angles, which corresponds to about a third of the port's thickness.

9.3.2. Proposed optimal single interface approximation for projection and reconstruction using the viewing volume

The model described in the previous section is only equivalent to the thick port model, if all object points have the same radial distance in the camera image. Otherwise, only an approximation of the optics of the port is possible with one interface, due to the angular dependency of Δd. For this common case with multiple different radial distances, an optimal solution with respect to a certain cost function can be defined. Agrawal et al. [73] defined this cost function by the reprojection error of the calibration target used.

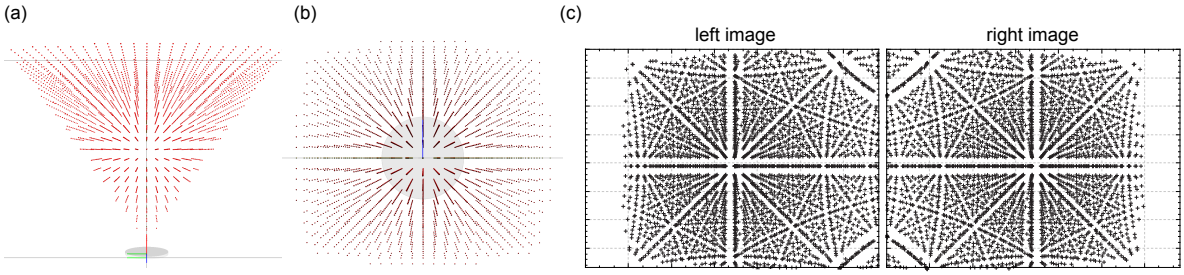

Figure 9.3.: Discretised viewing volume of a flat port underwater stereo camera and the corresponding stereo image pair: (a) Viewing volume, top view; (b) viewing volume, front view; (c) left and right camera image

However, this approach results in a solution that is only optimal for the points measured and, in addition to that, does not incorporate the 3D reconstruction error, which results in a different Δd as we will show.

We propose a solution that minimises the maximum 2D projection error $\hat{e}_{2d}$ and the maximum 3D reconstruction error $\hat{e}_{3d}$ for a given viewing volume V. It is based on a two-step calibration, in which at first the precise two interface model is calibrated to subsequently generate a data set for the second calibration step, in which the single interface distances Δd_{2d} and Δd_{3d} for optimal 2D refractive forward projection and optimal 3D reconstruction are determined.

Since the viewing volume V contains an infinite number of points and, thus, is not numerically processable, we simplify V by a three-dimensional 5 mm point grid, which is limited to an expected underwater visibility of 1 m as shown in Fig. 9.3. In comparison to the viewing volume of usual stereo cameras, the viewing volume of an underwater camera with a flat port and rectified images is deformed. The viewing volume is shrunk and bounded by four outward curved surfaces. A similar effect is observable in the two-dimensional camera images (see Fig. 9.3).

The optimal interface shifts Δd_{2d} and Δd_{3d} to minimise the maximum 2D projection and 3D reconstruction errors, $\hat{e}_{2d}$ and $\hat{e}_{3d}$, are defined as follows:

$$\Delta d_{2d}^{*}(t,V) = \underset{\Delta d \in \mathbb{R}}{\arg\min} \underbrace{\max_{\boldsymbol{p} \in V,\, i \in \{l,r\}} e_{2d}(\boldsymbol{p}, \Delta d, t, i)}_{\hat{e}_{2d}}, \tag{9.5}$$

$$\Delta d_{3d}^{*}(t,V) = \underset{\Delta d \in \mathbb{R}}{\arg\min} \underbrace{\max_{\boldsymbol{p} \in V} e_{3d}(\boldsymbol{p}, \Delta d, t)}_{\hat{e}_{3d}}\,. \tag{9.6}$$

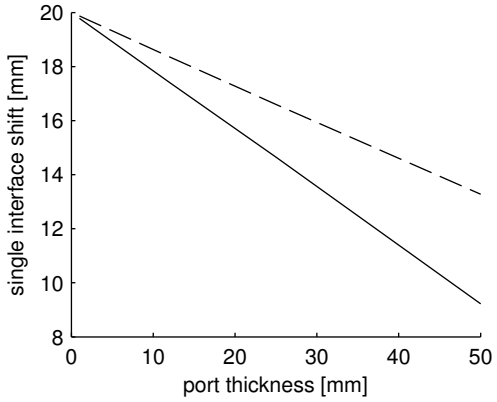

Figure 9.4.: Single interface approximation: The optimal interface shifts Δd_{2d} (solid line) and Δd_{3d} (dashed line) for different port thicknesses to minimise the maximum 2D projection and 3D reconstruction error, $\hat{e}_{2d}$ and $\hat{e}_{3d}$, in an approximated refractive stereo viewing volume. Due to the non-negligible difference between Δd_{2d} and Δd_{3d}, a separate calibration of $\hat{e}_{2d}$ and $\hat{e}_{3d}$ is recommended.

For the projection and reconstruction errors of a single point $\boldsymbol{p}$ we have

$$e_{2d}(\boldsymbol{p}, \Delta d, t, i) = |f_{rf}(\boldsymbol{p}, \Delta d, t' = 0, i) - \boldsymbol{u}_i|, \tag{9.7}$$

$$e_{3d}(\boldsymbol{p}, \Delta d, t) = |f_{rr}(\boldsymbol{u}_l, \boldsymbol{u}_r, \Delta d, t' = 0) - \boldsymbol{p}|, \tag{9.8}$$

where f_{rf} denotes the refractive forward projection (see chapter 5), f_{rr} the refractive reconstruction (see section 6.2), and where $\boldsymbol{u}_i = f_{rf}(\boldsymbol{p}, t, i)$ with $i \in \{l, r\}$ describes the correct image point of $\boldsymbol{p}$ in the left or right camera, respectively (see chapter 5). In summary, the proposed method determines the optimal interface shift for 2D projection and the optimal interface shift for 3D reconstruction using the viewing volume of the stereo camera created from the accurate thick port model.

9.4. Results

For our wide-angle underwater camera setup, the optimal interface shifts, $\Delta d_{2d}(t)$ and $\Delta d_{3d}(t)$, are depicted in Fig. 9.4 as a function of the port thicknesses t (Table 2.1, refractive indices as in 9.3.1). $\Delta d_{2d}(t)$ and $\Delta d_{3d}(t)$ are approximately linear for $0 \leq t < 5\,\text{cm}$, have different negative slopes and intersect at $t = 0$. The difference between the functions increases with t and results for our setup in a difference of about 1 mm at $t = 12\,\text{mm}$. This shows that the two interface shifts for optimal refractive forward

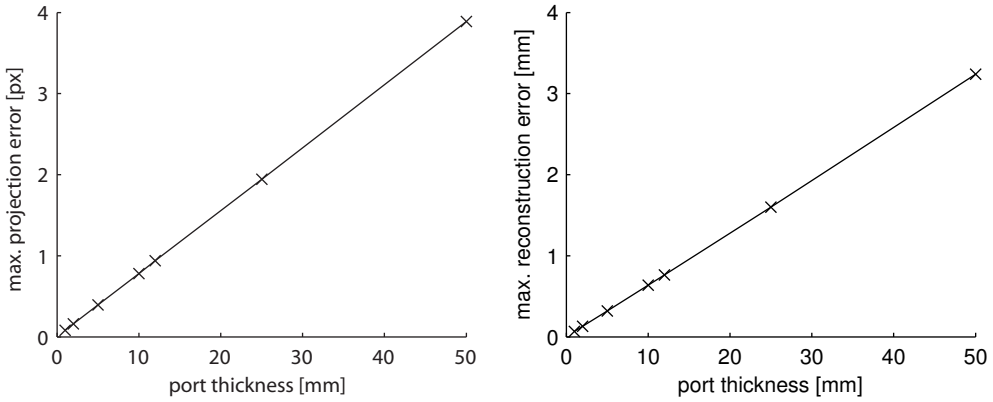

Figure 9.5.: Single interface approximation: The maximum projection error (left) and the maximum reconstruction error (right) of a stereo camera as a function of the port thickness

projection and optimal reconstruction need to be calibrated separately.

Finally, we will further investigate the accuracies in 2D point projection and 3D reconstruction achievable by single interface approximation. Based on the previous calibrations of the optimal interface shifts, the maximum 2D projection and 3D reconstruction errors, $\hat{e}_{2d}$ and $\hat{e}_{3d}$, are depicted in Fig. 9.5 with respect to the port thickness t. $\hat{e}_{2d}(t)$ and $\hat{e}_{3d}(t)$ grow with t and indicate a linear behaviour in the considered domain of 0 to 5 cm. For our setup, we obtain a maximum projection error of $\hat{e}_{2d} \approx 0.9$ px and a maximum reconstruction error $\hat{e}_{3d} \approx 0.8$ mm, which corresponds to 0.78 px/cm or to $\sim$6 % of the thickness of the port, respectively. Assuming a 2D feature point detection accuracy of $\sim$0.1 px, the results show that single interface approximation is measurable.

9.5. Conclusion

In this chapter, the projection and reconstruction performance of the single interface approximation for wide-angle thick, flat port underwater cameras was analysed. Additionally, a method to determine the optimal interface shifts for accurate projection and reconstruction was proposed using the stereo viewing volume of the thick port refractive camera model. Our research showed that a separate calibration of the interface shift for the projection and reconstruction is necessary to improve accuracy. However, the remaining errors of the single interface approximation are still measurable. For that reason, we recommend the use of the more accurate thick port refractive underwater camera model for wide-angle underwater stereo cameras.

10. Underwater surface image restoration

10.1. Introduction

Wide-angle underwater image acquisition through the flat port of an underwater housing distorts images with significant chromatic aberration and pincushion shaped image distortion. In this section, we propose two methods to minimise chromatic aberration and correct image distortion in large field of view, short baseline stereo, close-up, high resolution surfaces images. The proposed surface image restoration techniques are developed for the high quality texture mapping of reconstructed 3D ship hulls.

10.2. Minimisation of chromatic aberration

10.2.1. Background

Chromatic aberration is the result of wavelength dependent refraction of optical systems and consists of axial and lateral chromatic aberration. Axial chromatic aberration describes the effect when light beams of different wavelengths are not focused in a common image plane, whereas lateral chromatic aberration refers to the wavelength dependent radial shift in focus with respect to the optical axis [76]. Lateral chromatic aberration causes a different geometric distortion or magnification, respectively, of the colour channels, whereas axial chromatic aberration causes wavelength dependent blurring. Both axial and lateral chromatic aberration are visible as colour fringes at the edges of objects. Axial chromatic aberration can occur in the entire image, whereas lateral chromatic aberration increases with radial distance.

Different approaches are known to partially minimise chromatic aberration in underwater images. Chromatic aberration can be reduced by a combination of several positive and negative lenses that cause an overlapping of the focal points of light with different

wavelengths. An example of this is the Ivanoff corrector [62, 63] from 1956, which is mounted in front of an in-air lens. The corrector can partially correct chromatic aberration, but depends on the index of refraction of water, which can significantly vary as shown in section 6.2. For high quality correction of chromatic aberration, Ivanoff et al. suggested a multi lens systems, which increases complexity and significantly increases price. The Ivanoff corrector is also camera lens specific and only appropriate for medium field of views. Wakimoto [58] also criticised that the corrector was only designed on the basis of Gaussian optics, which describe the optical system only for small angles. As described in section 3.5, more advanced wet lens solutions of Nikon were built, but their production was stopped due to high manufacturing costs and small market size.

Chromatic aberration arises if the angle between the incident light ray and the normal of the interface is greater than zero. Assuming an ideal pinhole camera system and a precise alignment with the camera lens, dome ports can theoretically completely avoid chromatic aberration as the direction of the incident ray in water always coincides with the according surface normal. In practice, however, chromatic aberration still exists [58]. The fabrication and alignment is difficult, and in addition to that, dome ports significantly deform the virtual image resulting in reduced corner sharpness particularly visible at a large aperture [58, 61]. Considering chromatic aberration and corner sharpness, large dome radii are recommended, but are not suitable for the short baseline stereo we need.

For in-air photography, a wide range of image processing solutions to minimise chromatic aberration exists. These algorithms are able to minimise lateral [120, 121, 122, 123] and also axial chromatic aberration [121, 124]. The methods presented in [120, 121, 122, 123], minimise chromatic aberration by warping. Most authors [120, 121, 122] use the green colour channel as reference, as it has higher resolution and is often focused, and modified the red and blue colour channel to compensate for the wavelength dependent magnification or blurring effects. To determine the amount of warping and to calibrate the corresponding model parameters, single colour image registration is usually carried out. Boult and Wolberg [123] modelled chromatic aberration by interpolating the regions between image control points. Kaufmann and Ladstädter [122] and Mallon and Whelan [120] developed a chromatic aberration model, which is based on Brown's [60] standard lens distortion model with an additional first order term to incorporate the variation of the index of refraction. Kang [121] compensated magnification by a radial polynomial function and reduced defocus by a sharpening filter. The method of Chang et al. [124] differs from the warping methods, as it only corrects the visible artefacts locally at

edges. Chang et al. use a transient method to reduce colour blurring and a false colour filtering technique to compensate for the shift in colour. Due to the correction of only local characteristics, this algorithm only enhances the visual image appearance but does not recover the actual image geometry. In general, image processing-based solutions are not able to eliminate chromatic aberration completely. Since image sensors are sensitive to a finite number of narrow spectral bandwidths (RGB), correcting all wavelengths of the light spectrum is not possible. However, the advantage of image processing-based solutions lies in their cost efficiency and easy adaptability to various camera geometries.

In-air algorithms to minimise the effects of chromatic aberration are generally not designed for in-water application. Since the chromatic aberration of underwater cameras with a flat port varies with scene distance, the performance of in-air chromatic aberration techniques is limited. Changes in chromatic aberration due to scene depth cannot be modelled precisely for instance by polynomial functions as in prior research and so requires more complex models with a larger number of parameters.

We propose an alternative image processing method to reduce lateral chromatic aberration, which is based on the physics-based model of an underwater camera with a flat port. Our method requires a calibrated underwater camera including a calibration of the three major wavelengths and information about the scene depth, which is often already available by bundle adjustment or SLAM. Since our method is physics-based, the complexity to model and reduce chromatic aberration is limited. If the scene depth is known, our accurate method can be directly applied and so does not require additional time consuming image analysis.

10.2.2. Proposed image processing pipelines to minimise chromatic aberration

Our two proposed alternative image processing pipelines are illustrated in Fig. 10.1. They mainly differ in the final demosaicing step. The first pipeline is more efficient and completely based on bilinear demosaicing. The second pipeline is slower but allows the use of arbitrary more sophisticated demosaicing algorithms, which can significantly increase accuracy.

Demosaicing is necessary, as the majority of today's colour camera sensors uses a Bayer pattern for colour detection (see Fig. 10.2). In order to detect colour, 50 % of the pixels have a green, 25 % a red and 25 % a blue colour filter. Since each pixel can only detect one colour, missing colours need to be reconstructed by demosaicing.

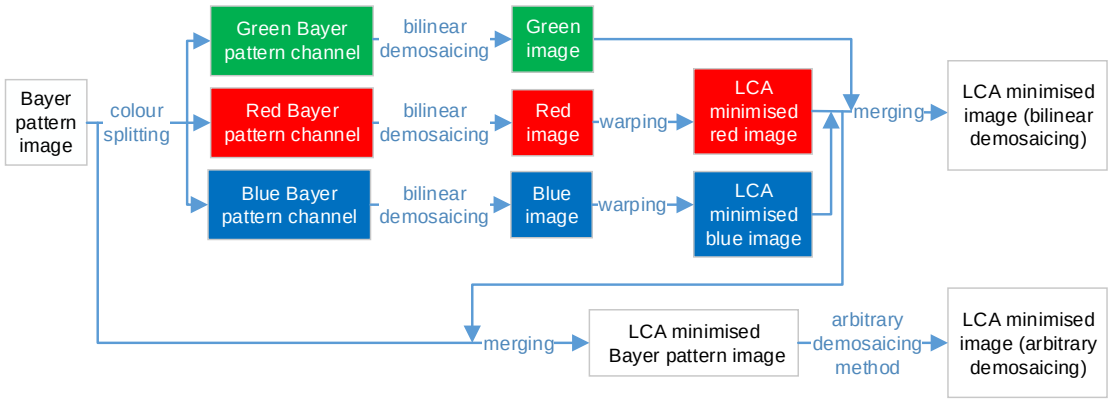

Figure 10.1.: The image processing pipeline to minimise lateral chromatic aberration: The upper part of the diagram shows the correction of chromatic aberration using bilinear demosaicing in all processing steps. The lower part of the diagram depicts an alternative way in which an intermediate Bayer pattern image is generated in order to use a more accurate demosaicing method in the final processing step.

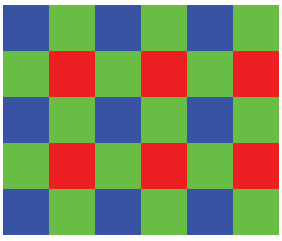

Figure 10.2.: A Bayer pattern filter as found in the majority of today's cameras. To enable the application of arbitrary demosaicing algorithms to generate the final chromatic aberration minimised image, we propose two alternative image processing pipelines.

The first image processing pipeline proposed was developed for efficient minimisation of chromatic aberration and efficient colour reconstruction and uses bilinear demosaicing in all steps. To correct chromatic aberration, the red, green and blue colour channels are split at first from the raw image and demosaiced separately by bilinear demosaicing. Subsequently, the image distortion of the red and blue colour channel with respect to the green colour channel is computed. This step is explained in more detail in section 10.2.3. The calculated distortion data is then used to minimise lateral chromatic aberration by warping the red and blue colour channel. In the last step, the demosaiced green and the corrected red and blue colour channel are merged to generate the final image with minimised lateral chromatic aberration.

The alternative image processing pipeline for arbitrary demosaicing algorithms differs only slightly. Instead of splitting, (bilinear) demosaicing and merging the green colour channel with the corrected red and blue colour channel, we propose the generation of an artificial Bayer pattern image from the original raw (Bayer pattern) image and the corrected red and blue channel. This method allows the application of arbitrary more accurate demosaicing techniques. As an example, we selected the DCB demosaicing algorithm of Góźdź [125] [126], as it successfully reduces zipper effects at edges, which would have been created by bilinear demosaicing. DCB was mainly developed for an attractive overall image look and does not use any post processing step, such as noise reduction, sharping or minimisation of chromatic aberration. DCB is one of those demosaicing algorithms, which exploit the correlation between the colour channels and take advantages of the higher resolution of the green channel for red and blue colour reconstruction.

10.2.3. Proposed pixel-wise calculation of chromatic aberration for the red and blue colour channel

To correct lateral chromatic aberration in the red and blue (demosaiced) colour channel by warping, the distortion of the channels with respect to the green reference channel need to be determined. The distortion describes the mapping from discrete coordinates $\boldsymbol{u}_g$ in the green colour channel to real image coordinates, $\boldsymbol{u}_r$ and $\boldsymbol{u}_b$, in the red or blue colour channel

$$\begin{aligned} f_{ca} :\ & \mathbb{N}^2 \to \mathbb{R}^2 \times \mathbb{R}^2, \\ & \boldsymbol{u}_g \mapsto (\boldsymbol{u}_r, \boldsymbol{u}_b). \end{aligned} \tag{10.1}$$

As previously described, lateral chromatic aberration in underwater cameras with a flat port is scene distance dependent. To compensate for lateral chromatic aberration, we use the physics-based model and information about the shape of the scene and its relative pose to the camera. To determine the corresponding scene distance for each image pixel, the first step in our method comprises the transformation of distorted to undistorted image coordinates. Using the green wavelength, the undistorted image points are subsequently back projected by refractive back projection to obtain the corresponding light rays in water. In the following step, the intersection points of the back projected light rays in water with the scene are calculated. Next, the intersection points are reprojected twice back onto the camera image using refractive forward projection with both the red and blue wavelength. The algorithm ends with a final transformation from undistorted to distorted image coordinates. The approach is summarised as follows:

$$\begin{aligned} f_{ca} : \boldsymbol{u}_r &= f_{dc}(f_{rf}(a, \lambda_r)), \\ \boldsymbol{u}_b &= f_{dc}(f_{rf}(a, \lambda_b)), \\ \boldsymbol{a} &= f_{irs}(f_{rb}(f_{uc}(\boldsymbol{u}_g), \lambda_g)),\ a \in \mathbb{R}^3. \end{aligned} \tag{10.2}$$

Here, f_{uc} and f_{dc} denote the transformation from distorted to undistorted coordinates and vice versa. f_{rb} describes the refractive back projection with the green wavelength λ_g, f_{irs} the intersection of the back projected ray with the scene and f_{rf} the refractive forward projection with the red or green wavelength λ_r or λ_b.

For an efficient calculation of f_{ca}, several lookup tables are used. The use of only one lookup table is not possible, as the scene distance changes and the scene points $\boldsymbol{p}_i$ need to be computed repeatedly frame by frame with f_{irs}. The mapping from distorted image coordinates to corresponding light rays in water is calculated using the first lookup table

$$\begin{aligned} f_{lut} : \mathbb{N}^2 &\rightarrow \mathbb{R}^3 \times \mathbb{R}^3, \\ \boldsymbol{u}_g &\mapsto f_{rb}(f_{uc}(\boldsymbol{u}_g), \lambda_g). \end{aligned} \tag{10.3}$$

For the refractive forward projection of scene points, Maas's lookup table-based refractive forward projection method $f_{rf,maas}$ is deployed (see chapter 5), which works in the plane of incidence and requires the corresponding coordinate transformations, f_{poi} and f_{poi}^{-1}, to and from the plane of incidence (5.3.1) and a subsequent perspective projection

f_{pf}.

$$f_{rf} : \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}^2,$$
$$(\boldsymbol{p}, \lambda) \mapsto f_{pf}(f_{poi}^{-1}(f_{rf,maas}(f_{poi}(\boldsymbol{p}), \lambda))). \tag{10.4}$$

The third and last lookup table is part of the coordinate transformation from undistorted to distorted image coordinates

$$f_{dc} : \mathbb{R}^2 \to \mathbb{R}^2 \tag{10.5}$$

and used in combination with bilinear interpolation.

10.2.4. Results

To evaluate the performance of our physics-based lateral chromatic aberration minimisation method, our underwater camera and a checkerboard were placed in a water tank as shown in Fig. 10.3. In this case, the scene captured consisted of a simple 3D plane whose relative pose to the camera was determined by checkerboard corner extraction and sparse bundle adjustment (chapter 7). The distance between the camera and the checkerboard was approximately 70 cm.

The left column of Fig. 10.4 shows a single underwater image, which was processed in the traditional way using only bilinear demosaicing and no correction of lateral chromatic aberration. As can be seen, the image is noisy and clearly exhibits blue and yellow colour fringes around the white squares of the checkerboard as a result of chromatic aberration.

To reduce the noise, we took advantage of the static setup and averaged 100 shots. The noise reduction step was applied to make the reduction in colour fringing in the final image better visible. The noise reduced image was subsequently processed by our proposed chromatic aberration minimisation approach and DCB demosaicing.

The final image is shown in the middle column of Fig. 10.4. As a result of our lateral chromatic aberration minimisation approach, the image shows a strong reduction of colour fringes resulting in higher edge sharpness. Another difference is observable in the form of a green margin in the outer regions of the image. This is unfortunately caused by the limited implementation of the transformation f_{uc} from distorted/rectified to undistorted/rectified image coordinates of the camera's manufacturer. The implementation of f_{uc} is not bijective, that is, a complete mapping from coordinates in the distorted image space to coordinates in the undistorted image space was not provided.

Figure 10.3.: Underwater experiment to evaluate the performance in minimising chromatic aberration and correcting refractive image distortion: The checkerboard and the underwater camera inside a water tank are shown.

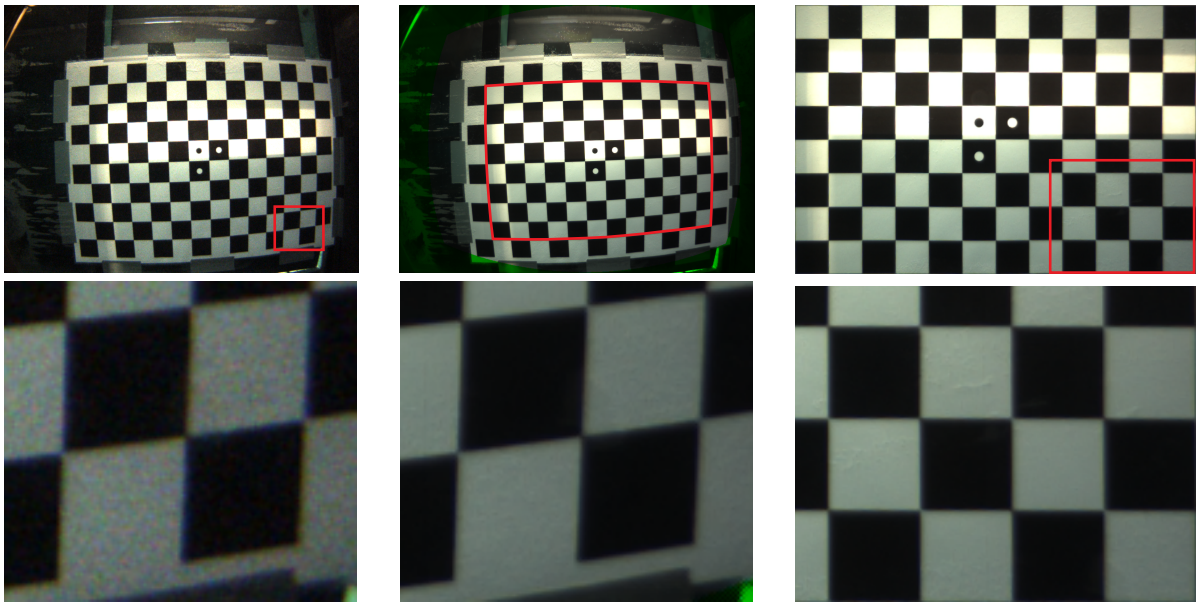

Figure 10.4.: Image processing steps for surface image restoration: The left column shows a single noisy underwater image of a checkerboard with visible chromatic aberration (yellow and blue fringes) and distortion. The middle column shows the noise reduced and lateral chromatic aberration corrected image obtained after several processing steps. The right column shows the extracted, undistorted inner checkerboard section (see red frame in in the middle column) with a resolution of 1 px/mm.

Table 10.1.: Execution times for the minimisation of chromatic aberration of a 0.6 MP image using bilinear demosaicing

	Time
Channel splitting (red + green + blue)	1.4 ms
Bilinear demosaicing (red + green + blue)	3.1 ms
CA calculation (red + blue)	69 ms
Warping (red + blue)	5.9 ms
Merging	1.4 ms
Total	82 ms

Table 10.2.: Execution times for the minimisation of chromatic aberration of a 0.6 MP image using DCB demosaicing in the final image processing step

	Time
Channel splitting (red + blue)	0.9 ms
Bilinear demosaicing (red + blue)	2.2 ms
CA calculation (red + blue)	69 ms
Warping (red + blue)	5.9 ms
Artifical Bayer pattern image	0.3 ms
DCB demosaicing	176 ms
Total	255 ms

Thus, LCA minimisation in the outer image region of the red and blue channels was not possible.

The execution times of our bilinear and DCB demosaicing-based lateral chromatic aberration minimisation methods are compiled in Table 10.1 and Table 10.2. The execution times were obtained on a Core i7-2600 CPU using multi-threading and efficient software libraries [118, 127] as far as possible. The complete calculation of chromatic aberration of a 1024×768 px image required the processing of ∼786 thousand pixels, whereas actually only ∼612 thousand pixels passed the complete f_{ca} computation step due to the missing bijectivity of f_{uc} as described before. The computation of chromatic aberration required only ∼69 ms and underlines the efficiency of our approach. Although the execution time for a more complex scene could increase with the increased computational effort to determine the scene depth by f_{irs}, in our experiment, f_{irs} just represents the calculation of the intersection point between a line and a plane. The total execution time of our algorithm proves that pixel-wise, accurate physics-based minimisation of lateral chromatic aberration can be realised efficiently. In our experiments, the bilin-

ear demosaicing-based method was executed in 82 ms and the DCB demosaicing-based approach in 255 ms. This shows that our method is also appropriate for deployment in real-time applications.

10.3. Correction of refractive image distortion for texture mapping

10.3.1. Introduction

Distortion is an obvious effect immediately visible in our underwater camera images. The distortion is the result of two interfering distortions, namely pincushion and barrel distortion, caused by the refraction of light at the flat port of the underwater housing and by the ultra wide-angle lens of the camera. In our setup, barrel distortion dominates and is visible as a magnification, which is greater in the image centre (see Fig. 10.4). For efficient stereo matching, correction of lens distortion in combination with image rectification is performed in in-air stereo vision [116]. For underwater cameras with a flat port, deploying distortion correction in the context of epipolar constrained feature matching for 3D reconstruction is unreliable, as the distortion introduced by refraction is not constant and changes with distance (see chapter 4). An application for refractive undistortion is surface image reconstruction for texture mapping in photorealistic 3D reconstruction.

In underwater 3D reconstruction, polygon meshes are a common 3D data visualisation technique. For enhancing appearance, the polygon surfaces is usually texturized. Vertex colouring and UV mapping represent two common texturing techniques.

In vertex colouring, the texture colour is generated by interpolating the colour information of the polygon's vertices. Vertex colouring is often used in the context of point clouds. Although not explicitly mentioned, we guess that vertex colouring was used by Kang et al. [82] and Jordt [78]. The reconstructed model of Kang et al. is based on a point cloud, which is relatively sparse. Thereby, the reconstructed textures look quite distorted and blurry. Kang et al.'s work uses the accurate physic-based refractive camera model. However, the thickness of the port and its tilt are still simplified to allow an easier calculation of refractive forward projection. As shown in section 6.4, the approximation of the interface tilt causes large projection errors. If, in contrast to that, the point cloud is dense enough with respect to the polygon size, vertex colouring can generate textures, which are very close to their perspectively correct original. This

was shown by Jordt [78] whose work is also based on the physically correct underwater camera model, but includes also the tilt and thickness of the port.

If the point cloud is sparse, more photorealistic textures can be realised by UV texturing as presented in [51, 128, 49, 110, 129, 111]. These works applied the perspective camera model in combination with the distortion model to perspectively project camera images onto the mesh. However, the perspective camera model and lens distortion model cannot imitate refractive distortion accurately, as refractive distortion is distance dependent (see chapter 4). The distortion of textures becomes particularly visible in close range reconstruction with sparse point clouds and large polygons where the scene distance differs from the average calibration distance.

Our proposed method uses refractive forward projection to create distortion free UV maps. It is based on the precise physics-based model, which also includes the port's thickness and tilt. In this way, it is possible to create accurate, high resolution textures for sparse surface meshes. In contrast to prior research, we demonstrate that refractive forward projection is also suitable for the projection of a large number of object points. The real-time capability of our method is demonstrated by generating an undistorted 0.3 MP image on a standard CPU in less than 15 ms.

10.3.2. Related literature

The first publications on the correction of refractive distortion in images captured through the wall of a water tank or through the port of an underwater housing appeared about 30 years ago. In 1986, Fryer and Fraser [64] modelled the refractive distortion of analogue cameras by adding a linear term to the standard radial lens distortion model. The coefficient of the new term depended on the focal length of the perspective camera model and an additional variable focal length correction term. Thus, the model was able to correct the distortion of flat scenes located at a constant distance to the camera.

In 1999 and 2000, Kwon [130] tried to compensate for refractive object plane deformation by linear methods, and reported that the non-linear effects of refractive distortion do not depend on the image coordinates alone.

Within the last years, several authors investigated the capability of a modified perspective camera model to approximate refractive image distortion. Kunz and Singh [59] showed that the perspective camera model in combination with an adapted focal length and with the standard distortion model is inaccurate for refractive undistortion and particularly if the port is tilted. For a single interface, Treibitz et al. [68, 71] showed that an adapted perspective camera model in combination with an extended radial dis-

tortion model is only accurate if the effective centre of projection is exactly located on the refractive interface. Only in this case, all incident light rays are focused in a single point.

10.3.3. Proposed method

There are several ways to map the image of an object onto the object's surface. Based on the physics-based refractive camera model, the image-to-texture mapping is possible by refractive back projection or refractive forward projection. Refractive back projection of image points onto the mesh surface has the advantage of being more efficient but results in a point structure, which is difficult to assign to the regular pixel grid of the texture, as the back projected light rays create an unstructured point pattern on the mesh surface. An alternative way circumventing this issue is based on the forward projection of the surface texture points onto the camera image. Using a sufficiently high texture resolution to avoid undersampling of the distorted image, the intensity of each texture pixel can be obtained by interpolating the intensity values of the corresponding pixels in the distorted image. This approach is similar to the warping procedure deployed in the chromatic aberration minimisation method (section 10.2). The mapping from object points on the polygon to image points of the camera is described by the following equation:

$$\begin{aligned} f_{ud} : \mathbb{R}^3 &\rightarrow \mathbb{R}^2, \\ \boldsymbol{p} &\mapsto \boldsymbol{u}, \\ \boldsymbol{u} &= f_{dc}(f_{rf}(\boldsymbol{p})). \end{aligned} \tag{10.6}$$

Similar to section 10.2, $\boldsymbol{u}$ denotes the distorted point coordinates, f_{dc} the transformation from undistorted to distorted camera image coordinate, f_{rf} the refractive forward projection and $\boldsymbol{p}$ an object point on the polygon corresponding to a pixel in the texture. To obtain the complete UV texture map, f_{ud} is applied for all corresponding texture points.

10.3.4. Results

Our experimental setup is very similar to the lateral chromatic aberration experiment (10.2.4) and depicted in Fig. 10.4. Again a checkerboard was attached to the wall of a water tank with an underwater camera placed in front of it. A checkerboard was chosen,

Table 10.3.: Execution times for the correction of distortion of a 0.3 MP surface image for texture mapping

	Time
Refractive distortion calculation	12.4 ms
Warping	1.7 ms
Total	14.2 ms

as it enables an accurate evaluation of the distortion correction performance of the algorithm proposed. As shown in Fig. 10.4, a large checkerboard region was selected in order to demonstrate the accuracy and efficiency of our algorithm. The selected checkerboard region had a size of $67.1 \times 44.8\,\mathrm{cm}^2$ and was sampled with a sufficiently high resolution of 1 px/mm resulting in a texture size of 671 × 448 px. For better image quality, the texture calculation was applied on the preprocessed image from section 10.2 for which we firstly averaged 100 camera images of the checkerboard, then minimised chromatic aberration using our proposed method (see section 10.2). The pose of the checkerboard was determined again by bundle adjustment. The generated texture is shown in Fig. 10.4. The high quality of the undistorted image is evidenced by checkerboard lines being parallel and right-angled, which we check by an image viewing program. Again, the lookup table-based refractive forward projection method of Maas [1] was used (see chapter 5). In contrast to prior research our method is efficient enough to be deployed in real-time applications. The execution time of our proposed method to process $671 \times 448\,\mathrm{px} = 0.3\,\mathrm{MP}$ on a Core i7-2600 CPU required only 14.2 ms.

10.4. Conclusion

In this chapter, we have proposed two novel methods to restore the surface images captured by a wide-angle flat port underwater stereo camera. On the one hand, a method was demonstrated to accurately minimise the lateral chromatic aberration, which is caused by the refraction of light at the port and is visible as distinct colour fringes at sharp edges in the image. On the other hand, a method to correct the strong image distortion was presented, which is a result of the superposition of the barrel distortion of the lens and the pincushion distortion of the refraction at the port. The minimisation of the chromatic aberration and the correction of the barrel distortion were developed for the high quality texture mapping of reconstructed 3D ship hulls. Despite the accu-

rate pixel-wise image processing, the suitability of both methods for real-time texture mapping was demonstrated.

11. Pseudo rectified images and epipolar curves for underwater correspondence

11.1. Introduction

Classic image undistortion is used to correct the distortion by the camera lens and the misalignment of the camera's imager. Image undistortion maps straight lines in object space to straight lines in image space.

Classic image rectification is a method in stereo vision for facilitated stereo correspondence search. Here, object points are imaged onto the same horizontal straight epipolar line.

Image undistortion and image rectification were originally developed for in-air imaging where the classic perspective camera model applies. These in-air techniques were also deployed in numerous underwater vision prior research, such as ship hull inspection [131, 49], dam inspection [50], underwater archaeology [109] [51] [132] or natural underwater environment mapping [110] [112, 113] [133], [111]. In flat refractive underwater imaging however, where the object space is imaged through one or more parallel refractive interfaces (port, window, water surface), the imaging process is not perspective, any longer. The distortion caused by the refraction of light at the interfaces is not constant but depends on the scene distance. The perspective camera model or the more general single viewpoint model [68, 71], respectively, become invalid. Consequently, classic in-air camera techniques applied in flat refractive underwater imaging are inaccurate.

To avoid inaccuracies in flat refractive imaging, an accurate physics-based model is used. This model works with image coordinates of in-air undistorted/rectified camera images. Due to refraction, these images appear strongly distorted underwater, particularly with ultra wide-angle camera lenses. As a result, image distortion diminishes the performance of feature matching algorithms, such as SIFT [119] or speeded up robust

features (SURF) [134], which incorporate larger image regions in the feature detection and description process.

We will show how image distortion of underwater cameras with one or more parallel flat interfaces can be minimised for facilitated feature matching using accurate physics-based pseudo rectified images. Furthermore, we will propose an approach for computing epipolar curves for accurate and real-time epipolar constrained underwater stereo correspondence search.

11.2. Related literature

Stereo correspondence search in flat underwater imaging using the physics-based model has been addressed by only few authors. In 2005, Ferreira et al. [135] used a first order linearisation of Snell's law to obtain epipolar lines and to rectify distorted underwater images. Using this first order approximation, also referred to as Gaussian optics, is only valid for small angles, though.

In the theoretical work of 2009, Chari and Sturm [69] mathematically demonstrated in analogy to epipolar lines in perspective projection that epipolar lines do not exist for the single refractive interface case. In their work, it was shown that the image points of an object point seen by two cameras are actually related by a quartic curve.

Also Gedge et al. [70] computed refractive epipolar curves using a physics-based model of a single refractive interface. In their method from 2011, firstly, the corresponding ray in water of each image point was computed by refractive back projection. Next, the intersection points of the ray in water, with planes at different distances, were determined, and subsequently projected onto the second camera image by refractive forward projection. In a final step, the epipolar curves were modelled by piecewise linear approximation using the projected sampling points. Gedge et al. used a stereo camera, which is very similar to our model and mainly differs in the longer focal length (4 mm vs. 2 mm) and lower resolution (640×480 px vs. 1024×768 px). The authors measured a mean approximated epipolar constraint error of 0.54 px, whereas we achieved in our more challenging setup with cubic epipolar curves a mean error of only $\sim$0.19 px. Another disadvantage of Gedge et al.'s method lies in the time-consuming computation of the epipolar lines, which is not desirable in real-time stereo. Gedge et al. also used in-air rectified images for feature matching. As mentioned before, this leads to strongly distorted images with wide-angle underwater cameras.

In their work on dense underwater stereo from 2013, Jordt-Sedlazeck et al. [77] used a

refractive plane sweep method for image undistortion to avoid refractive forward projection claimed to be computationally costly. Instead of using refractive forward projection, image points were back projected onto multiple 3D hypothesis planes using a lookup table for the image point-to-in-water ray mapping. A warped image of the reference view for each hypothesis plane was obtained by computing the intersections of the back projected rays with the hypothesis planes. As the irregular pattern of the intersection points did not coincide with the image rasterization on the hypothesis plane, a polygon from the intersection points was spanned over the plane and the rasterization performed by a graphics processing unit (GPU). In two subsequent plane sweep steps, a dense stereo map was obtained using stereo correspondence search. In our opinion, this approach is too complicated and based on wrong assumptions. Jordt-Sedlazeck et al. did not only claim that refractive forward projection is computationally too expensive, but also that dense stereo with epipolar lines for guided correspondence search is not feasible. Our research also differs to the work of Jordt-Sedlazeck et al. in the camera lens used. Whereas Jordt-Sedlazeck et al. used a normal lens with narrow angles, our research is based on ultra wide-angle camera lenses, which cause even larger image distortions.

In some aspects, our proposed methods use parts of the ideas of the authors introduced. Our method is similar to the method of Gedge et al. in the way that we compute epipolar curves by back and forward projection. But instead of projecting the object points on the rays in water back to the rectified camera image, we projected them on a pseudo rectified image plane located in water. This is done to minimise the strong distortion in wide-angle underwater camera images to facilitate feature matching. The pseudo rectified image planes are similar to the hypothesis planes of Jordt-Sedlazeck et al.'s. In contrast to the image warping of Jordt-Sedlazeck et al.'s, we do not compute the mapping from camera to pseudo rectified image coordinates, but instead compute the opposite path from pseudo rectified to distorted camera image coordinates. This avoids the complicated rasterization process. Even though we do not explicitly demonstrate dense stereo, we show that using our pseudo rectified images, the epipolar curves can be precisely modelled in a compact form by a polynomial function and efficiently restored by a lookup table. In this way, and also due to the use of the more accurate thick port physics-based model, the computation of epipolar curves is more precise and more time and memory efficient than the method of Gedge et al.'s.

Our work is structured as follows. In section 11.3.1 and section 11.3.2, the calculation of the proposed pseudo rectified images with according coordinate system transformations is explained. Section 11.3.3 reports on the proposed algorithm to calculate refractive

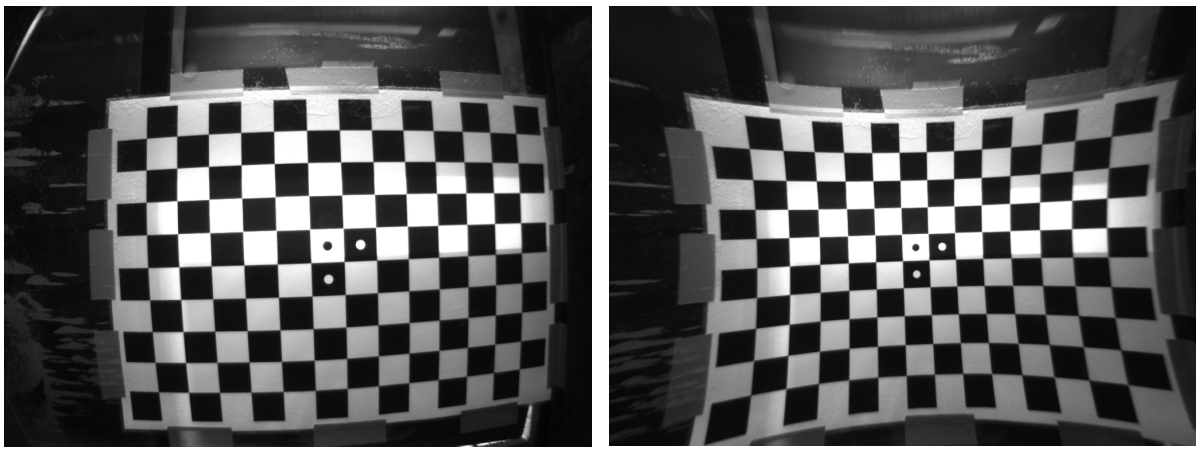

Figure 11.1.: Image distortion: (a) barrel distortion in the original underwater camera image, (b) strong pincushion distortion in the (in-air) rectified image

epipolar curves. In section 11.4, we present the results from real pseudo rectified images captured in an underwater tank and the according refractive epipolar curves.

11.3. Proposed methods

11.3.1. Proposed pseudo rectified image coordinate system

To understand the idea of pseudo rectified images, an overview of the possible image distortions in flat refractive underwater imaging is given. At first, consider the left raw image of a wide-angle underwater camera. As shown in Fig. 11.1a, clear barrel distortion is observable. As illustrated in Fig. 11.2a, this barrel distortion is mainly the result of the superposition of the pincushion distortion caused by the refraction of light at the interfaces of the port and of the barrel distortion of the ultra wide-angle camera lens, which dominates in this case.

Image undistortion minimises the effected of lens distortion. In in-air imaging, image undistortion makes the resulting image appear as if it were captured in front of the lens as shown in Fig. 11.2b. Flat refractive underwater imaging is also based on image coordinates given in the in-air undistorted (rectified) image coordinate system. However, as shown in Fig. 11.1b, in-air distorted (rectified) underwater images captured through a flat port and an ultra wide-angle lens lead to strong image distortion.

Image distortion deteriorates the performance of correspondence search algorithms, which are usually designed for undistorted images. The performance of correspondence

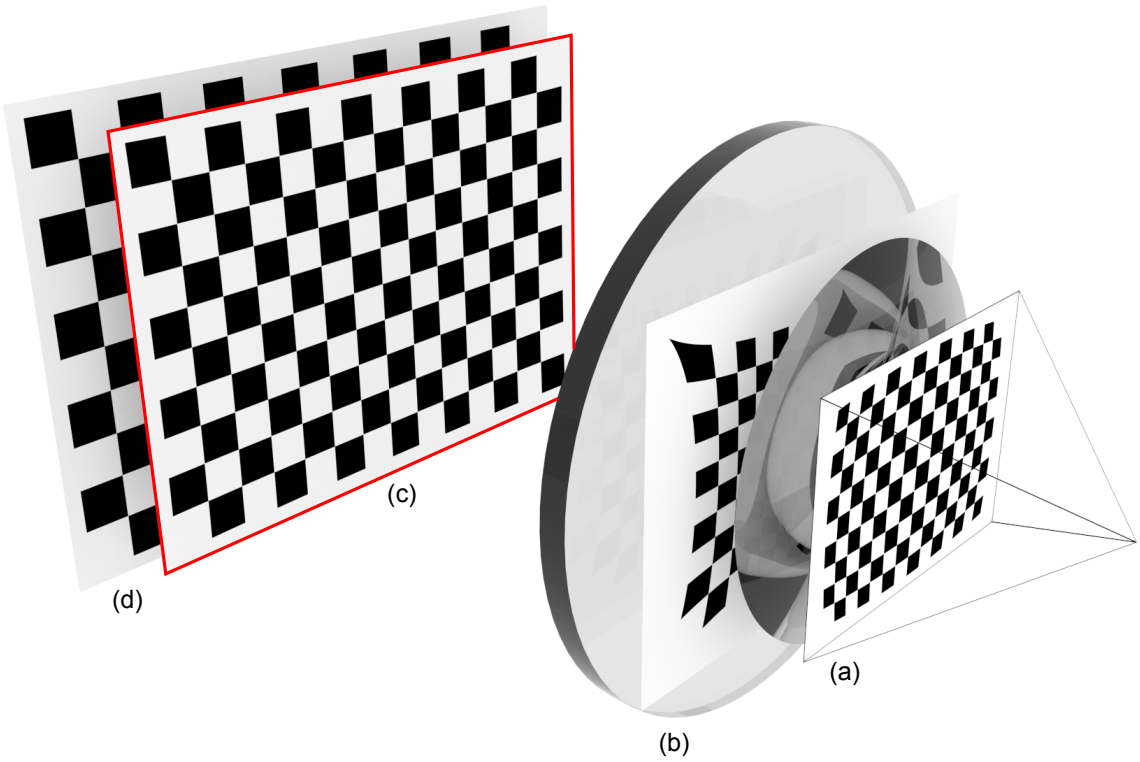

Figure 11.2.: Distortions in flat underwater imaging and the idea of a pseudo undistorted (rectified) image: (a) barrel distortion in the original underwater camera image seen through a flat port and a camera lens, (b) strong pincushion distortion in the (in-air) undistorted (rectified) image seen through a flat port, (c) less distortion in the pseudo undistorted (rectified) image reconstructed in water for facilitated correspondence search, (d) the actual image on a flat object

search algorithms, which analyse the environment around interest points, is diminished the more the distortion around two interest points differs.

There are several ways to reduce the image distortion for facilitated correspondence search. One way is to follow the traditional way and to use a general distortion model, such as the standard lens distortion model of Brown [60] for example. The resulting image would be sufficient for correspondence search, but requires an additional calibration of the parameters of the distortion model. Instead, our image undistortion takes advantage of the physics-based model, which already completely describes the optical geometry of the underwater camera. The advantage of this approach is that no additional calibration is required and the image distortion can be directly derived.

In our stereo camera application, we do not only minimise the distortion of both camera images, but create two images, which look similar to rectified camera images. Rectified images have the property that object points are projected onto the same horizontal (epipolar) line. Rectified images are not rotated, are displaced only vertically to each other and image object points with the same focal length. Because of these properties, we approximated in-air rectified images for flat refractive stereo. Due to the distance dependency of refractive image distortion and if the scene distance is unknown, image rectification in flat refractive imagery is possible only partially. Therefore, we refer to these images as pseudo rectified images. Pseudo rectified images mimic rectified images precisely if the scene is flat and the image plane of the pseudo rectified images coincides with the actual scene plane. Also in the more common case with scenes of different depths, pseudo image undistortion/rectification reduces image distortion significantly. As we will also show in section 11.3.3, pseudo rectified images enable the modelling of refractive epipolar curves with polynomials of low degree. Our proposed methods are based on a preliminary refractive calibration of the pose of the underwater housing's port, which was described in chapter 7.

Let $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z} \in \mathbb{R}^3$ be the axes of the stereo camera frame, where $\boldsymbol{x}$ is parallel to the optical axis of the rectified camera images and points towards the image plane of the pseudo rectified image, and where $\boldsymbol{y}$ is parallel to the camera's baseline and points to the left as shown in Fig. 11.3. The pseudo rectified image plane is placed at the distance $x = d_p$ parallel to the (in-air) rectified images. Additionally, let $\boldsymbol{u}_{ul,l}$, $\boldsymbol{u}_{ur,l}$, $\boldsymbol{u}_{ll,l}$, $\boldsymbol{u}_{lr,l} \in \mathbb{R}^2$ be the corners of the left (in-air) rectified image and $\boldsymbol{u}_{ul,r}$, $\boldsymbol{u}_{ur,r}$, $\boldsymbol{u}_{ll,r}$, $\boldsymbol{u}_{lr,r} \in \mathbb{R}^2$ be the image corners of the right (in-air) rectified image. The computation of the coordinate systems of the pseudo rectified images corresponds to a mapping f_{prcs} from the eight image corners $\boldsymbol{u}_{ul,l}, \ldots, \boldsymbol{u}_{lr,r}$ to the origins $\boldsymbol{O}_l$ and $\boldsymbol{O}_r$ of the left and right

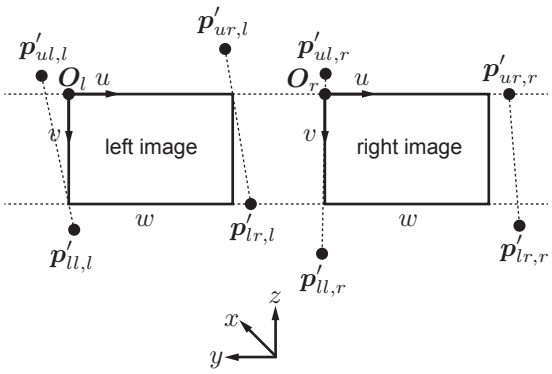

Figure 11.3.: Coordinate systems of the pseudo rectified camera images and of the stereo camera

pseudo rectified images (Fig. 11.3)

$$\begin{aligned} f_{prcs} &: \mathbb{R}^{2\times 8} \times \mathbb{R} \to \mathbb{R}^{3\times 2} \times \mathbb{R}, \\ (\boldsymbol{u}_{ul,l}, \ldots, \boldsymbol{u}_{lr,r}, w) &\mapsto \boldsymbol{O}_l \times \boldsymbol{O}_r \times s, \end{aligned} \tag{11.1}$$

where w denotes the desired width in pixels of the pseudo rectified images and results in the scaling factor s to scale coordinates in the stereo camera frame to coordinates in the pseudo rectified image frame.

The computation of the pseudo rectified image coordinate system starts with the refractive back projection f_{rb} of $\boldsymbol{u}_{ul,l}, \ldots, \boldsymbol{u}_{lr,r}$ onto the image plane of the pseudo rectified image. The resulting eight intersection points $\boldsymbol{p}_{ul,l}, \ldots, \boldsymbol{p}_{lr,r} \in \mathbb{R}^3$ of the back projected rays from the left and right camera with the pseudo rectified image plane (Fig. 11.3) are given by

$$\boldsymbol{p}_{ij,k} = f_{irp}(f_{rb}(\boldsymbol{u}_{ij,k}), k, d_p), \quad i \in \{u, l\},\ j, k \in \{l, r\}. \tag{11.2}$$

The points form two quadrilaterals with a non-rectangular shape, which need to be cropped to two rectangles. According to in-air rectified images, they should

1. have the same size, which should be as large as possible,
2. have the same orientation,
3. are only displaced side-ways,
4. and are as close as possible to each other to maximise image overlapping for stereo correspondence search.

To reduce additionally the rotation between the (in-air) rectified and the pseudo rec-

tified images, let $\boldsymbol{u} = -\boldsymbol{y}$ and $\boldsymbol{v} = -\boldsymbol{z}$ be the horizontal and vertical axes of the pseudo rectified image frames (Fig. 11.3). The offsets $\boldsymbol{O}_l$ and $\boldsymbol{O}_r$ of the pseudo rectified image frames can be now computed in several steps.

First, the coordinates of the eight points $\boldsymbol{p}_{ul,l}, \ldots, \boldsymbol{p}_{lr,r}$ given in the 3D stereo camera frame are mapped to the 2D coordinates $\boldsymbol{p}'_{ul,l}, \ldots, \boldsymbol{p}'_{lr,r}$ in the pseudo rectified image plane by

$$\boldsymbol{p}'_{ij,k} = (-p_{ij,k,y}, -p_{ij,k,z})^T, \tag{11.3}$$

where $p_{ij,k,y}$ and $p_{ij,k,z}$ denote the y- and z-coordinate of $\boldsymbol{p}_{ij,k}$.

Since the pseudo rectified images are displaced only vertically and are not rotated with respect to each other, they shared the same upper and lower horizontal edge with the vertical coordinates v_u and v_l. These edges are defined by the lowest of the upper four points and the highest of the lower four points, respectively. More specifically, we have

$$\begin{aligned} v_u &= \max_{j,k \in \{l,r\}} (p'_{uj,k,v}), \\ v_l &= \min_{j,k \in \{l,r\}} (p'_{lj,k,v}). \end{aligned} \tag{11.4}$$

Thus, the upper and lower horizontal lines defining the upper and lower edges of the images are given by

$$\begin{aligned} l_u &= ((0, v_u)^T, (1, 0)^T), \\ l_l &= ((0, v_l)^T, (1, 0)^T). \end{aligned} \tag{11.5}$$

To maximise the overlapping of both images, the left image should be as far as possible right and the right image as far as possible left. This is achieved by considering the two inner lines $l_{i,l}$ and $l_{i,r}$ defined by the points $\boldsymbol{p}'_{ur,l}$ and $\boldsymbol{p}'_{lr,l}$, and $\boldsymbol{p}'_{ul,r}$ and $\boldsymbol{p}'_{ll,r}$, as shown in Fig. 11.3. Using these lines, the horizontal coordinates $u_{i,l}$ and $u_{i,r}$ of the inner edges of the left and right images are defined by

$$\begin{aligned} u_{i,l} &= \min(f_{ill}(l_u, l_{i,l})_u, f_{ill}(l_l, l_{i,l})_u), \\ u_{i,r} &= \max(f_{ill}(l_u, l_{i,r})_u, f_{ill}(l_l, l_{i,r})_u), \end{aligned} \tag{11.6}$$

where $f_{ill}(l_1, l_2)_u$ denotes the horizontal coordinate of the intersection point of the lines l_1 and l_2.

To maximise the image width, the vertical coordinates $u_{o,l}$ and $u_{o,r}$ of the potential outer image edges are computed similarly by

$$\begin{aligned} u_{o,l} &= \max(f_{ill}(l_u, l_{o,l})_u, f_{ill}(l_l, l_{o,l})_u), \\ u_{o,r} &= \min(f_{ill}(l_u, l_{o,r})_u, f_{ill}(l_l, l_{o,r})_u), \end{aligned} \tag{11.7}$$

where $l_{o,l}$ and $l_{o,r}$ denote the outer lines of the left and right image defined by the points $\boldsymbol{p}'_{ul,l}$ and $\boldsymbol{p}'_{ll,l}$, and $\boldsymbol{p}'_{ur,r}$ and $\boldsymbol{p}'_{lr,r}$, respectively (see Fig. 11.3).

Knowing the position of the inner image edges and the position of the potential outer image edges, the common width w' of both images is determine by

$$w' = \min(u_{i,l} - u_{o,l}, u_{o,r} - u_{i,r}). \tag{11.8}$$

Finally, the origins, $\boldsymbol{O}_l$ and $\boldsymbol{O}_r$, of the pseudo rectified images and the scaling s from stereo camera to pseudo rectified image coordinates is given by

$$\boldsymbol{O}_l = \begin{pmatrix} p_{ul,l,x} \\ w - u_{i,l} \\ -v_u \end{pmatrix}, \quad \boldsymbol{O}_r = \begin{pmatrix} p_{ul,l,x} \\ -u_{i,r} \\ -v_u \end{pmatrix}, \quad s = w/w'. \tag{11.9}$$

11.3.2. Proposed coordinate transformations

Using a pseudo rectified image, our proposed method requires three coordinate mappings in total, that is:

1. A mapping f_{pr} from (in-air) rectified image coordinates (Fig. 11.2b) to pseudo rectified image coordinates (Fig. 11.2c), which is necessary if points in the refractive camera image are required in the pseudo rectified image (sampling points of the pseudo epipolar curve, see Fig. 11.4; bundle adjustment; SLAM)
2. A mapping f_{pr}^{-1} from pseudo rectified image coordinates (Fig. 11.2c) to (in-air) rectified image coordinates (Fig. 11.2b) to transform the coordinates of detected feature points for the refractive camera model
3. And a mapping f_{pr2dc} from pseudo rectified image coordinates (Fig. 11.2c) to distorted image coordinates (Fig. 11.2a) for the image warping.

The mapping from (in-air) rectified image coordinates $\boldsymbol{u}$ to pseudo rectified image coordinates $\boldsymbol{u}_{pr}$ of the left or right camera

$$\begin{aligned} f_{pr} : \mathbb{R}^2 \times \{l, r\} &\to \mathbb{R}^2, \\ (\boldsymbol{u}, i) &\mapsto \boldsymbol{u}_{pr}, \end{aligned} \tag{11.10}$$

is defined by the refractive back projection f_{rb} of $\boldsymbol{u}$, the computation of the intersection point of the back projected ray with the pseudo rectified image plane using f_{irp},

$$\boldsymbol{p} = f_{irp}(f_{rb}(\boldsymbol{u}, i), d_p) \tag{11.11}$$

and a subsequent coordinate transformation including offset adjustment and scaling based on $\boldsymbol{O}_l$ or $\boldsymbol{O}_r$ and s

$$\boldsymbol{u}_{pr} = s \begin{pmatrix} O_{i,y} - p_y \\ O_{i,z} - p_z \end{pmatrix}. \tag{11.12}$$

The inverse coordinate transformation from pseudo rectified image coordinates to (in-air) rectified image coordinates

$$\begin{aligned} f_{pr}^{-1} : \mathbb{R}^2 \times \{l, r\} &\to \mathbb{R}^2, \\ (\boldsymbol{u}_{pr}, i) &\mapsto \boldsymbol{u} \end{aligned} \tag{11.13}$$

is similar. First, $\boldsymbol{u}_{pr}$ is transformed to the point $\boldsymbol{p}$ in stereo camera coordinates using

$$\boldsymbol{p} = -\frac{1}{s} \begin{pmatrix} 0 \\ u_{pr} \\ v_{pr} \end{pmatrix} + \boldsymbol{O}_i, \tag{11.14}$$

and subsequently projected by refractive forward projection f_{rf} onto the (in-air) rectified left or right image

$$\boldsymbol{u} = f_{rf}(\boldsymbol{p}, i). \tag{11.15}$$

The coordinate transformation from discrete pseudo rectified image coordinates to (in-air) distorted image coordinates

$$\begin{aligned} f_{pr2dc} : \mathbb{N}^2 \times \{l, r\} &\rightarrow \mathbb{R}^2, \\ (\boldsymbol{u}_{pr}, i) &\mapsto \boldsymbol{u}_{dc} \end{aligned} \tag{11.16}$$

is based on the coordinate transformation f_{pr}^{-1} from pseudo rectified image coordinates to (in-air) rectified coordinates and includes an additional mapping f_{dc} from (in-air) rectified coordinates to distorted coordinates, as describes by

$$\boldsymbol{u}_{dc} = f_{dc}(f_{pr}^{-1}(\boldsymbol{u}_{pr}, i)). \tag{11.17}$$

For an efficient computation of all three mappings, we use lookup tables in combination with bilinear interpolation for f_{pr} and f_{pr}^{-1}.

11.3.3. Proposed modelling of epipolar curves in pseudo rectified images

Our motivation to use epipolar lines lies in the additional constraint for feature matching in pseudo rectified camera images. Using pseudo rectified images, each image point in the left camera image has a corresponding epipolar curve in the right camera image showing where the left image points is seen in the right camera image depending on the distance of the 3D object point.

Since we are interested in real-time feature matching of pseudo rectified images, our work focuses on an epipolar curve model, which allows a fast check of the epipolar constraint. Given the image points $\boldsymbol{u}_{pr,l}$ and $\boldsymbol{u}_{pr,r}$ in the left and right image, we use a model, which enables a fast computation of the epipolar constraint error, that is, the vertical distance d between the epipolar curve and $\boldsymbol{u}_{pr,r}$,

$$\begin{aligned} f_{ece} : \mathbb{R}^2 \times \mathbb{R}^2 &\rightarrow \mathbb{R}, \\ (\boldsymbol{u}_{pr,l}, \boldsymbol{u}_{pr,r}) &\mapsto d. \end{aligned} \tag{11.18}$$

To achieve this goal, we model each epipolar curve in the right image for each discrete

pixel $\boldsymbol{u}_{pr,l,i} \in \mathbb{N}^2$ in the left image by the polynomial function

$$\begin{aligned} f_{pf} &: \mathbb{R} \to \mathbb{R}, \\ u_{pr,r} &\mapsto v_{pr,r}, \\ v_{pr,r} &= c_0 + c_1 u_{pr,r} + c_2 u_{pr,r}^2 + \cdots + c_n u_{pr,r}^n, \\ u_{pr,r} &\leq u_{pr,l}, \end{aligned} \tag{11.19}$$

which enables a fast calculation of the vertical coordinate $v_{pr,r}$ of the epipolar curve as a function of the horizontal curve coordinate $u_{pr,r}$. An advantage of the polynomial representation f_{pf} of the epipolar curve is the compact form enabling the use of a lookup table to store the coefficients $c_0, \ldots, c_n$ for each discrete image point $\boldsymbol{u}_{pr,l,i}$. In combination with bilinear interpolation, epipolar curves can be computed efficiently for arbitrary real input values $\boldsymbol{u}_{pr,l}$. More details about bilinear interpolation can be found in section 5.5.2.

To further reduce the computation time of the approximated epipolar curves, we determine the smallest polynomial degree n, which still results in small approximation errors. We use least-squares curve fitting to calculate the coefficients c_{ji} of each curve. To obtain the sampling points $\boldsymbol{u}_{pr,r,1}, \ldots, \boldsymbol{u}_{pr,r,m}$ for each epipolar curve, each point $\boldsymbol{u}_{pr,l,i}$ of the left image is firstly back projected by refractive back projection onto planes with different distances $d_{p,1}, \ldots, d_{p,m}$, which are parallel to the camera image planes. Subsequently, the back projected points are projected onto the right camera image using refractive forward projection. Finally, the projected points are transformed from (in-air) rectified to pseudo rectified coordinates. In summary, we have

$$\boldsymbol{u}_{pr,r,j} = f_{pr}(f_{rf}(f_{irp}(f_{rb}(\boldsymbol{u}_{pr,l,i}, l), d_{p,j}), r), r), \quad j \in \{1, \ldots, m\}, \tag{11.20}$$

where f_{irp} denotes the intersection point of the back projected ray with a plane at the distance $d_{p,j}$. To obtain an approximately uniform distribution of the sampling points $\boldsymbol{u}_{pr,r,1}, \ldots, \boldsymbol{u}_{pr,r,m}$ along the $\boldsymbol{u}$-axis, we calculate the plane distances $d_{p,1}, \ldots, d_{p,m}$ by

$$d_{p,j} = \frac{(\hat{d}_p - \check{d}_p)}{j} + \check{d}_p, \tag{11.21}$$

where $\check{d}_p$ and $\hat{d}_p$ denote the minimum and maximum plane distance chosen. This formula compensates for the perspective-like distance dependency of the horizontal position of the sampling points.

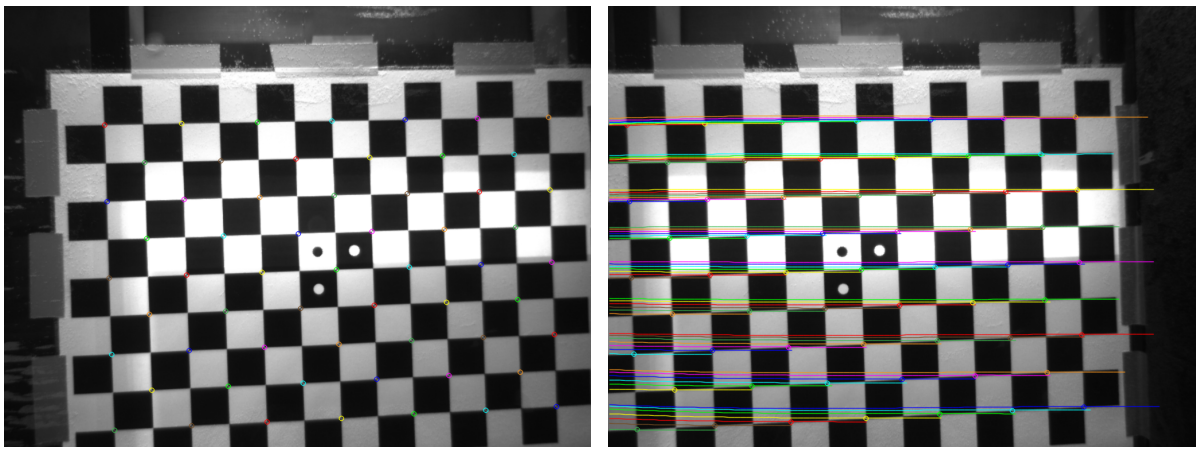

Figure 11.4.: Proposed pseudo rectified images and epipolar curves: (a) the left pseudo rectified camera image, (b) the corresponding right pseudo rectified camera image with epipolar curves

11.4. Results

The result of our proposed pseudo rectification algorithm applied on the left distorted camera image (Fig. 11.1a) is depicted in Fig. 11.4a. The pseudo rectified image plane was placed at the distance $d_p = 90\,\text{cm}$ whereas the actual plane distance amounted to $\sim$70 cm. As can be seen, the image distortion in the pseudo rectified image is significantly reduced in comparison to the original distorted images (Fig. 11.1). The pseudo undistortion or rectification, respectively, is comparable to the classic underwater image undistortion/rectification approach using the perspective camera model with an adapted focal length in combination with the standard lens distortion model. Unlike this classic method, our method is based on the physics-based model and thus allows accurate coordinate transformation from pseudo rectified image coordinates to distorted raw image coordinates for image warping and accurate coordinate transformation between pseudo rectified image coordinates and in (in-air) rectified image coordinates, which is required in flat refractive geometry.

The pseudo rectified right camera image together with some epipolar curves is shown in Fig. 11.4b. For the generation of the sampling points, we placed the closest plane 5 cm in front of the port and the farthest plane at a large distance of 4 m. Using $m = 100$ planes, each epipolar curve was generated in average by $\sim$45 sample points. The number of sampling points is smaller than m, as very close points and points in the left part of the left image are not always visible in the right image. To minimise the polynomial

degree, curve fitting with different polynomial degrees was applied and the maximum error between the curve and the sampling points measured. Using an epipolar curve of degree three, the maximum vertical fitting error $\hat{e}$ was less than 0.35 px. For higher order polynomial degrees, this maximum epipolar constrained error reduced to 0.12 px or 0.07 px for degree four or five, respectively.

To evaluate the accuracy of the cubic epipolar curve in real underwater images captured by our calibrated underwater camera, we used a checkerboard, detected the checkerboard corners in the left and right camera image (see Fig. 11.4), and computed the mean and maximum epipolar constraint error $\bar{e}$ and $\hat{e}$. Here, we measured for $\bar{e} \approx 0.19$ px and $\hat{e} \approx 0.73$ px.

Furthermore, we analysed the average execution time to calculate the epipolar constraint error $f_{ece}(\boldsymbol{u}_{pr,l}, \boldsymbol{u}_{pr,r})$. Therefore, we calculated for all 1024×768 epipolar curves, all possible function values in the defined discrete range of 1 to $u_{pr,l}$ pixels and measured the overall execution time. Using all eight threads of our Intel Core i7-2600 CPU, the total computation time of all epipolar lines with an average width of 512 px amounted to $\sim$128 ms.

11.5. Conclusion

In this chapter, we showed that the refraction of light at the flat port of a wide-angle underwater camera leads to a strong pincushion distortion in the camera images originally rectified for the application in air. Since in-air rectified image coordinates are required in accurate physics-based flat refractive geometry, but the strong pincushion distortion can diminish correspondence search, we proposed a pseudo undistorted image with reduced distortion to facilitate correspondence search. This pseudo undistorted image results from a refractive back projection of the distorted image on a plane located in water, where the actual mapping for image warping is realised by refractive forward projection. The proposed pseudo undistorted image is not completely distortion free, but exhibits significantly less distortion than the in-air undistorted/unrectified image or the distorted raw camera image. To obtain two intermediate images for stereo vision, which are for example not rotated to each other and show enough overlapping for stereo matching, we additionally proposed so called pseudo rectified images and corresponding image space coordinate transformations. Based on these pseudo rectified images, we also proposed a way to accurately model epipolar curves for an epipolar constrained correspondence search. Whereas the refractive epipolar curve computation of Gedge

et al. [70] is time-consuming due to repeating refractive back and forward projection, we demonstrated how the epipolar curves can be modelled in a compact and time efficient form by polynomial functions from interpolated lookup table data. Computing for all 1024×768 discrete images points the corresponding epipolar curves and computing on each of these curves all discrete points with a positive disparity, could all be efficiently achieved in only $\sim$128 ms. The proposed theoretical methods were also evaluated in a water tank experiment using our wide-angle underwater short baseline stereo camera. A test with a checkerboard demonstrated the high accuracy of our proposed epipolar curves. Compared with the results of Gedge et al. [70] who used a camera with a lower resolution and normal lenses, even our less accurate cubic epipolar curves were more accurate and exhibited a 65 % lower mean epipolar constraint error of only $\sim$0.19 px.

12. Ship hull relative pose estimation

12.1. Introduction

In this chapter, we finally describe the wide-angle stereo vision-based ship hull relative pose estimation system for the real-time in-water inspection of ship hulls. The proposed pose estimation system is the first of its kind, which incorporates the accurate physics-based refractive camera model into a SLAM-based pose estimation framework. In spite of the computationally more demanding algorithms used, the pose estimation works in real-time and thus can be, for example, deployed for the autonomous navigation of an underwater vehicle inspecting a ship hull. This chapter also shows that the purely vision-based tracking of ship hulls is possible without the aid of any sonar based means [136].

12.2. Related literature

There is almost no prior research, which focuses on the ship hull relative pose estimation with a stereo camera. The work of Negahdaripour and Firoozfam [24] from 2006 is one of the few ship hull relative pose estimation projects, which used a stereo camera. Negahdaripour and Firoozfam's deployed an inaccurate perspective camera-based underwater camera model and estimated the pose of the camera with respect to a planar projection of the hull. The pose estimation was based on visual odometry and did not correct for consequential drift. The pose estimation accuracy of this stereo vision-based system was evaluated in a pool and a dry-dock. Here, the relative error in the estimated distance travelled was 2 %, which is about one order of magnitude worse than in our proposed method. Theses relative large inaccuracies also became visible in the strong distortion of the created photomosaics.

In 2007, Hogue et al. [51] used an underwater stereo camera, which was very similar to our model. The underwater camera was used for the reconstruction of a sunken barge, which also incorporated a visual odometry-based pose estimation. The inaccurate

perspective camera-based underwater camera model was used to reconstruct a dense point cloud. The point cloud points were tracked by the Kanade–Lucas–Tomasi (KLT) feature tracker [137] and cleaned from outliers by random sample consensus (RANSAC) [138, 139]. The relative pose between the reconstructed point clouds was then determined by Horn's absolute orientation algorithm [115] and subsequently refined by non-linear optimisation, but without using more accurate bundle adjustment. An evaluation of the accuracy of the pose estimation accuracy under water was not performed.

The PeRL group at the University of Michigan also focuses on the inspection of ship hulls, but with the major goal of detecting mine-line objects [49, 20]. The work is based on the VAN framework, which uses mainly a sonar-based DVL for the pose estimation. The navigation is augmented with a single monocular camera, which is primarily used for the correction of the arising drift with SLAM. Similar to our project, the camera detects landmarks on the ship hull and uses SLAM for the self-localisation of the robot and the mapping of the ship hull. In contrast to our research, the inaccurate perspective camera-based underwater camera model is used. In addition to that, the camera is not used for permanent high-frequency ship hull tracking, but runs with a low frequency and establishes only occasionally constraints between partially overlapping key frame images. The feature extraction is realised by SIFT [140] and executed on a GPU with a low frequency of 2-3 Hz. For the estimation of the robot pose a graph-based non-linear optimisation framework [141] is used.

12.3. Extended Kalman filter simultaneous localisation and mapping

SLAM [33, 34] was firstly mentioned in the 1980s and has become popular over the last two decades. SLAM enables a system with adequate perception and processing capabilities to localise itself in an initially unknown environment. In the context of SLAM, this system, which is often called robot, incrementally creates a map of the environment and uses this map for self-localisation. The advantage of SLAM to dead reckoning-based localisation system is the ability to correct arising drifts by considering the environment as a fixed reference system. Nowadays, SLAM can be found in various applications on land, in air and underwater. It is especially useful where no external positioning information, such as from the Global Positioning System (GPS), is available.

A great variety of SLAM solutions exits, where filtering-based SLAM and graph-based SLAM represent two major groups. Filtering-based approaches fuse the information

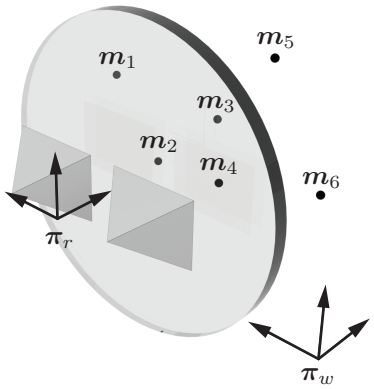

Figure 12.1.: Self-localisation with SLAM: In the self-localisation or pose estimation process, respectively, a robot starts from a world frame $\boldsymbol{\pi}_w$ and estimates its pose $\boldsymbol{\pi}_r$ using fixed landmarks $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_n$ of the environment.

of all measurements in probability distributions, whereas graph-based methods collect the measurement data and use it later for the accurate non-linear global estimation of large maps and the corresponding robot poses. In contrast to filtering based-techniques, graph-based methods can only process a subset of the available measurement data, but are able to handle wrong data associations between the measurements and the landmarks.

The extended Kalman filter (EKF) is a Bayesian filter, which estimates the state of a non-linear system with normally distributed noise [142]. The system state is estimated by fusing the information about the system behaviour and state related observations. In a repetitive approach, the system behaviour is predicted by a state transition model and subsequently corrected in an update step using real observations and the corresponding observation model.

12.3.1. State description

Our underwater SLAM system estimates the pose $\boldsymbol{\pi}_r$ of the stereo camera or robot, respectively, relative to the starting pose of the robot, which is used as the world frame $\boldsymbol{\pi}_w$ (see Fig. 12.1). The robot has six degrees of freedom (DOF), which are represented by a 13 elements state vector $\boldsymbol{x}_r$. It consists of

- ${}^w\boldsymbol{r}$ the 3-D position of the robot in the 3-D Euclidean space relative to the world frame,
- ${}^w\boldsymbol{q}$ the orientation of the robot relative to the world frame represented by an orientation quaternion,

${}^{w}\boldsymbol{v}$ the linear velocity of the robot in the world frame,
${}^{r}\boldsymbol{\omega}$ and the robot's angular velocity described in the robot frame.

In addition to the state of the robot, the EKF also maintains a map ${}^{w}\boldsymbol{m}^{T}$ of the observed 3D landmarks $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_n$ (see Fig. 12.1), which increases over time while moving through the environment. The complete system state is denoted by $\boldsymbol{x} = (\boldsymbol{x}_r^T \; {}^{w}\boldsymbol{m}^T)^T$.

The EKF is a Bayesian estimator and incorporates the probability distribution of $\boldsymbol{x}$ in the estimation process. The probability distribution of $\boldsymbol{x}$ is approximated by its mean value $\bar{\boldsymbol{x}}$ and the error covariance matrix $\boldsymbol{P}$

$$\bar{\boldsymbol{x}} = \begin{pmatrix} \bar{\boldsymbol{x}}_r \\ \bar{\boldsymbol{m}} \end{pmatrix}, \; \boldsymbol{P} = \begin{pmatrix} \boldsymbol{P}_{x_r x_r} & \boldsymbol{P}_{x_r m} \\ \boldsymbol{P}_{m x_r} & \boldsymbol{P}_{mm} \end{pmatrix}, \tag{12.1}$$

$$\bar{\boldsymbol{x}}_r = \begin{pmatrix} {}^{w}\bar{\boldsymbol{r}} \\ {}^{w}\bar{\boldsymbol{q}}_r \\ {}^{w}\bar{\boldsymbol{v}} \\ {}^{r}\bar{\boldsymbol{\omega}} \end{pmatrix}, \; \bar{\boldsymbol{m}} = \begin{pmatrix} {}^{w}\bar{\boldsymbol{m}}_1 \\ {}^{w}\bar{\boldsymbol{m}}_2 \\ \vdots \\ {}^{w}\bar{\boldsymbol{m}}_n \end{pmatrix}. \tag{12.2}$$

The elements on the diagonal of the covariance matrix represent the uncertainties of the estimated state variables, while the off-diagonal elements describe the correlations between the variables. The covariance matrix creates a fully connected correlation network between the landmarks and has a useful effect in the EKF update step (12.3.4), but is likewise one of the main limitations of EKF-SLAM.

As shown in Fig. 12.2, a part of the SLAM state $\boldsymbol{x}$ can be conveniently visualised by axes and ellipsoids, where the axes describe the estimated mean pose $\bar{\boldsymbol{\pi}}_r = ({}^{w}\bar{\boldsymbol{r}}^T, {}^{w}\bar{\boldsymbol{q}}_r^T)^T$ of the robot and the ellipsoids the estimated means and uncertainties $({}^{w}\bar{\boldsymbol{r}}, \boldsymbol{P}_{rr})$ and $({}^{w}\bar{\boldsymbol{m}}_i, \boldsymbol{P}_{m_i m_i})$ of the robot position or of the landmarks, respectively.

12.3.2. State transition model

The state transition model $\boldsymbol{f} : \boldsymbol{x}_{k-1} \mapsto \boldsymbol{x}_k$ specifies the time-dependent behaviour of the robot and the map from time step $k-1$ to time step k. Since the landmarks $\boldsymbol{m}$ are assumed to be fixed, $\boldsymbol{m}$ remains unchanged and only the motion $\boldsymbol{f}_r$ of the robot is modelled with respect to time

$$\boldsymbol{x}_k = \boldsymbol{f}(\boldsymbol{x}_{k-1}) = \begin{pmatrix} \boldsymbol{f}_r(\boldsymbol{x}_{r,k-1}) \\ \boldsymbol{m}_{k-1} \end{pmatrix}. \tag{12.3}$$

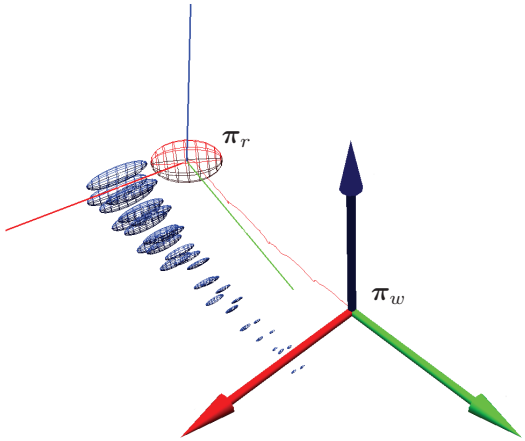

Figure 12.2.: Visualisation of the SLAM state: The estimated robot pose $\boldsymbol{\pi}_r$, the estimated landmark positions (blue ellipsoids), the robot trajectory (red line) together with the world frame $\boldsymbol{\pi}_r$ are shown.

We mainly use visual odometry $\boldsymbol{f}_{r,vo}$ to model the motion of the robot and use the constant velocity model $\boldsymbol{f}_{r,cv}$ of Davison et al. [143] as a fall-back solution to skip short periods without visual odometry

$$\boldsymbol{f}_r(\boldsymbol{x}_{r,k-1}) = \begin{cases} \boldsymbol{f}_{r,vo}(\boldsymbol{x}_{r,k-1},\, \Delta\boldsymbol{\pi}), & \text{if } \Delta\boldsymbol{\pi} \neq \emptyset, \\ \boldsymbol{f}_{r,cv}(\boldsymbol{x}_{r,k-1},\, \boldsymbol{N}), & \text{else} \end{cases} . \tag{12.4}$$

As described in more detail in section 12.4.2, visual odometry determines the relative pose $\Delta\boldsymbol{\pi}$ of the robot between two time steps from the observations of the stereo camera. $\Delta\boldsymbol{\pi}$ is represented by the translation ${}^r\Delta\boldsymbol{r}$ and the rotation quaternion ${}^r\Delta\boldsymbol{q}$, or more specifically, by its mean $\Delta\bar{\boldsymbol{\pi}}$ and its covariance matrix $\boldsymbol{P}_{\Delta\pi}$

$$\Delta\boldsymbol{\pi} = \begin{pmatrix} {}^r\Delta\boldsymbol{r} \\ {}^r\Delta\boldsymbol{q} \end{pmatrix} \sim \boldsymbol{\mathcal{N}}\left(\Delta\bar{\boldsymbol{\pi}},\, \boldsymbol{P}_{\Delta\pi}\right). \tag{12.5}$$

To incorporate visual odometry, we use the following robot motion model

$$\boldsymbol{f}_{r,vo}(\boldsymbol{x}_r,\, \Delta\boldsymbol{\pi}) = \begin{pmatrix} {}^w\boldsymbol{r} + {}^w\Delta\boldsymbol{r} \\ {}^w\boldsymbol{q}_r \oplus {}^r\Delta\boldsymbol{q} \ominus {}^w\boldsymbol{q}_r \\ {}^w\Delta\boldsymbol{r}/\Delta t \\ f_\omega({}^r\Delta\boldsymbol{q})/\Delta t \end{pmatrix},$$
$${}^w\Delta\boldsymbol{r} = {}^w\boldsymbol{q}_r \oplus {}^r\Delta\boldsymbol{r}, \tag{12.6}$$

where f_ω denotes the conversion of a rotation quaternion to an angle-axis vector.

For the rare case without visual odometry, we deploy the constant linear and angular velocity robot motion model with additive Gaussian noise of Davison et al. [143]

$$\boldsymbol{f}_{r,cv}(\boldsymbol{x}_r,\, \boldsymbol{N}) = \begin{pmatrix} {}^w\boldsymbol{r} + ({}^w\boldsymbol{v} + {}^w\boldsymbol{V})\,\Delta t \\ {}^w\boldsymbol{q}_r \oplus f_q\left(({}^r\boldsymbol{\omega} + {}^r\boldsymbol{\Omega})\,\Delta t\right) \\ {}^w\boldsymbol{v} + {}^w\boldsymbol{V} \\ {}^r\boldsymbol{\omega} + {}^r\boldsymbol{\Omega} \end{pmatrix}, \qquad \boldsymbol{N} = \begin{pmatrix} {}^w\boldsymbol{V} \\ {}^r\boldsymbol{\Omega} \end{pmatrix} \sim \boldsymbol{\mathcal{N}}\left(\boldsymbol{0}_{6,1},\, \operatorname{diag}\begin{pmatrix} \boldsymbol{\sigma}_V^2 \\ \boldsymbol{\sigma}_\Omega^2 \end{pmatrix}\right), \tag{12.7}$$

where f_q denotes the conversion from an angle-axis vector to a rotation quaternion. The effect of this model manifests in a smooth robot motion, if the visual odometry is not available and only few or no landmarks are visible.

12.3.3. Observation model

The observation model relates the system state $\boldsymbol{x}$ to expected observations $\tilde{\boldsymbol{z}}$. Thus, in the EKF update step (12.3.4), a correction of the system state can be carried out by comparing $\tilde{\boldsymbol{z}}$ with the actual observations.

In terms of our stereo-vision SLAM system, this means that a description of the expected image points $\tilde{\boldsymbol{z}}_1, \ldots, \tilde{\boldsymbol{z}}_n$ in the left and the right camera image, with $\tilde{\boldsymbol{z}}_{\boldsymbol{i}} = (u_{i,l}, v_{i,l}, u_{i,r}, v_{i,r})^T$, for all landmarks ${}^w\boldsymbol{m}$ from the current robot pose ${}^w\boldsymbol{\pi}_r = ({}^w\boldsymbol{r}^T\ {}^w\boldsymbol{q}_r^T)^T$ is needed

$$\boldsymbol{h}\left({}^w\boldsymbol{\pi}_r,\, {}^w\boldsymbol{m},\, \boldsymbol{W}\right) = \tilde{\boldsymbol{z}} = (\tilde{\boldsymbol{z}}_1, \ldots, \tilde{\boldsymbol{z}}_n)^T, \tag{12.8}$$

where $\boldsymbol{W}$ denotes additive, normal distributed observation noise

$$\boldsymbol{W} \sim \boldsymbol{\mathcal{N}}\left(\boldsymbol{0}_{4,1},\, \boldsymbol{U}\right),\ \boldsymbol{U} = \operatorname{diag}\begin{pmatrix} \boldsymbol{\sigma}_{uv}^2 \\ \boldsymbol{\sigma}_{uv}^2 \end{pmatrix}. \tag{12.9}$$

Based on our research in chapter 5, we used the fast lookup table-based refractive forward

projection method to compute the expected observations by

$$\tilde{\boldsymbol{z}}_i = \begin{pmatrix} u_{i,l} \\ v_{i,l} \\ u_{i,r} \\ v_{i,r} \end{pmatrix} = \begin{pmatrix} f_{rf}({}^r\boldsymbol{m}_i, l) \\ f_{rf}({}^r\boldsymbol{m}_i, r) \end{pmatrix} + \boldsymbol{W},$$
$${}^r\boldsymbol{m}_i = \ominus {}^w\boldsymbol{\pi}_r \oplus {}^w\boldsymbol{m}_i. \tag{12.10}$$

12.3.4. State prediction and update step

The state prediction and update step constitute the essential functionality of the extended Kalman filter. In the state prediction step, the new probability distribution of the system state $\boldsymbol{x}_{k|k-1}$ after the time Δt is estimated. Based on the previously estimated state mean $\bar{\boldsymbol{x}}_{k-1|k-1}$ the transition model is applied to estimate the new state mean

$$\bar{\boldsymbol{x}}_{k|k-1} = \boldsymbol{f}\left(\bar{\boldsymbol{x}}_{k-1|k-1},\, \boldsymbol{N} = \boldsymbol{0}\right). \tag{12.11}$$

In the same step, the new covariance matrix of the system state is estimated by a linear transformation of the previously estimated covariance matrix and an increase in the state's uncertainty by the drift of the visual odometry or the motion noise, respectively

$$\boldsymbol{P}_{k|k-1} = \frac{\delta \boldsymbol{f}}{\delta \boldsymbol{x}} \boldsymbol{P}_{k-1|k-1} \frac{\delta \boldsymbol{f}}{\delta \boldsymbol{x}}^T + \boldsymbol{M},$$
$$\boldsymbol{M} = \begin{cases} \frac{\delta \boldsymbol{f}}{\Delta \boldsymbol{\pi}} \boldsymbol{P}_{\Delta\pi} \frac{\delta \boldsymbol{f}}{\Delta \boldsymbol{\pi}}^T, & \text{if } \Delta\boldsymbol{\pi} \neq \emptyset, \\ \frac{\delta \boldsymbol{f}}{\delta \boldsymbol{N}} \operatorname{diag} \begin{pmatrix} \boldsymbol{\sigma}_V^2 \\ \boldsymbol{\sigma}_\Omega^2 \end{pmatrix} \frac{\delta \boldsymbol{f}}{\delta \boldsymbol{N}}^T, & \text{else} \end{cases}. \tag{12.12}$$

The update step of the extended Kalman filter fuses the information of the predicted system state with the information of the observations to a more accurate state estimate. Here, the observations are weighted depending on the degree of new information. The update step is described by the following well-known equations:

$$\boldsymbol{S}_k = \frac{\delta \boldsymbol{h}}{\delta \boldsymbol{x}} \boldsymbol{P}_{k|k-1} \frac{\delta \boldsymbol{h}}{\delta \boldsymbol{x}}^T + \operatorname{diag}\left(\left(\boldsymbol{\sigma}_{uv}^2{}^T \dots \boldsymbol{\sigma}_{uv}^2{}^T\right)^T\right), \tag{12.13}$$

$$\boldsymbol{K}_k = \boldsymbol{P}_{k|k-1} \frac{\delta \boldsymbol{h}}{\delta \boldsymbol{x}}^T \boldsymbol{S}_k^{-1}, \tag{12.14}$$

$$\bar{\boldsymbol{x}}_{k|k} = \bar{\boldsymbol{x}}_{k|k-1} + \boldsymbol{K}_k \cdot \left(\boldsymbol{z}_k - \boldsymbol{h}\left(\bar{\boldsymbol{x}}_{k|k-1},\, \boldsymbol{W} = \boldsymbol{0}\right)\right), \tag{12.15}$$

$$\boldsymbol{P}_{k|k} = \left(\boldsymbol{I} - \boldsymbol{K}_k \frac{\delta \boldsymbol{h}}{\delta \boldsymbol{x}}\right) \boldsymbol{P}_{k|k-1}. \tag{12.16}$$

Equations (12.15) and (12.16) show the final update of the estimated mean and covariance matrix.

Note, that (12.15) and (12.16) are the only places where the landmarks are modified. Since the state prediction step does not change the map, and the update step improves the state estimate, the accuracy of the estimated map incrementally increases.

12.3.5. Inverse observation model and map expansion

To expand the SLAM map with a new landmark ${}^{w}\bar{\boldsymbol{m}}_i$, ${}^{w}\bar{\boldsymbol{m}}_i$ needs to be reconstructed from the corresponding observation $\boldsymbol{z}_i$ of the stereo camera. This reconstruction is described by the inverse observation model

$$\begin{aligned} \boldsymbol{g}\left({}^{w}\boldsymbol{\pi}_r,\, \boldsymbol{z}_i,\, \boldsymbol{W}\right) &= {}^{w}\boldsymbol{m}_i = {}^{w}\boldsymbol{\pi}_r \oplus {}^{r}\boldsymbol{m}_i, \\ {}^{r}\boldsymbol{m}_i &= f_{rr}(\boldsymbol{z}_i + \boldsymbol{W}), \end{aligned} \tag{12.17}$$

where $\boldsymbol{z}_i$ is noisy and is firstly corrected by the random variable $\boldsymbol{W}_i$, subsequently reconstructed by the proposed refractive reconstruction method f_{rr} (see section 6.2), and finally transformed to the world coordinate system using the current robot pose ${}^{w}\boldsymbol{\pi}_r$.

Consequently, for the new expanded mean landmark map ${}^{w}\bar{\boldsymbol{m}}_{k,new}$ and the corresponding new covariance matrix $\boldsymbol{P}_{k|k,new}$ [144, 145], we have

$$\begin{aligned} {}^{w}\bar{\boldsymbol{m}}_{k,new} &= ({}^{w}\bar{\boldsymbol{m}}_{k|k}^{T}\ \boldsymbol{g}({}^{w}\bar{\boldsymbol{\pi}}_{r,k|k},\, \boldsymbol{z}_i,\, \boldsymbol{W}=0)^{T})^{T}, \\ \boldsymbol{P}_{k|k,new} &= \begin{pmatrix} \boldsymbol{P}_{x_r x_r,k|k} & \boldsymbol{P}_{x_r m,k|k} & \boldsymbol{P}_{x_r x_r,k|k}\frac{\delta \boldsymbol{g}}{\delta \boldsymbol{x}_r}^{T} \\ \boldsymbol{P}_{m x_r,k|k} & \boldsymbol{P}_{mm,k|k} & \boldsymbol{P}_{m x_r,k|k}\frac{\delta \boldsymbol{g}}{\delta \boldsymbol{x}_r}^{T} \\ \frac{\delta \boldsymbol{g}}{\delta \boldsymbol{x}_r}\boldsymbol{P}_{x_r x_r,k|k} & \frac{\delta \boldsymbol{g}}{\delta \boldsymbol{x}_r}\boldsymbol{P}_{x_r m,k|k} & \boldsymbol{A} \end{pmatrix}, \\ \boldsymbol{A} &= \frac{\delta \boldsymbol{g}}{\delta \boldsymbol{x}_r}\boldsymbol{P}_{x_r x_r,k|k}\frac{\delta \boldsymbol{g}}{\delta \boldsymbol{x}_r}^{T} + \frac{\delta \boldsymbol{g}}{\delta \boldsymbol{z}_i}\boldsymbol{U}\frac{\delta \boldsymbol{g}}{\delta \boldsymbol{z}_i}^{T}. \end{aligned} \tag{12.18}$$

12.3.6. Discussion of the advantages and disadvantages of EKF-SLAM

EKF-SLAM has some advantages, which make it interesting for real-time self-localisation and mapping. Due to the sensor data fusion, no sensor data needs to be accumulated allowing the processing of large amounts of data. Since there is a correlation network between all landmarks, the correction process does not only affect directly visible landmarks but the whole map including their correlations [33]. Furthermore, the EKF models

the uncertainties in the system state by normal distributions, which are represented in a compact form by their mean and their covariance, and so can be stored and processed very efficiently.

On the other hand, the probability distribution of the system state is not always normally distributed. Moreover, the maintenance of a fully correlated network leads to a computational complexity in the update step of $O(n^2)$, which limits the maximum number of landmarks to be processed. A drawback of the EKF also exists in the linearisation of non-linear models (Jacobians) and the modelling of probability distributions by normal distributions. Both approximations lead to inaccuracies in the non-linear transition model of the robot and the non-linear observation model. These are particularly visible, if the uncertainty in the orientation of the robot is large and affects all correlated state variables. Due to the data fusion, these non-linear errors cannot be corrected, afterwards. The same applies to errors by wrong data associations.

However, as our results in section 12.5 show, the uncertainties in the system state remain small and wrong data associations practically do not occur due to a robust outlier detection.

12.4. Image processing

12.4.1. From distorted raw images to correspondences for visual odometry and SLAM

We deploy SURF [134] to establish correspondences in real-time for stereo vision, visual odometry and SLAM. SURF describes and detects highly distinctive feature points, which are invariant to scale, rotation, translation, brightness and contrast. Therefore, SURF is used for the tracking of points on ship hulls. SURF is based on SIFT [140], but uses integral images and Haar-wavelets for higher speed.

Since the SURF feature descriptor is based on local gradient histograms, and so is not invariant to strong image distortion, we pre-process the raw images of the stereo camera and convert them to our proposed pseudo-rectified images (see chapter 11).

Since the refractive camera model requires point coordinates in in-air rectified images, the coordinates of the detected SURF key points in the pseudo rectified images are transformed to in-air rectified image coordinates before they are used in SLAM or visual odometry, respectively. For this purpose, the proposed functions f_{pr} and f_{pr}^{-1} from section 11.3.2 are used, but for brevity, not explicitly mentioned anywhere else in this

chapter.

Using another result of our research (see chapter 6), we used only the green demosaiced colour channel to reduce chromatic aberration and to increase the accuracy in feature detection. The basic structure of the image processing pipeline is depicted in Fig. 12.3.

SURF is used to establish correspondences for stereo vision, for visual odometry, the EKF update step and the EKF map expansion. To this end, correspondences are established between the left and right camera image (stereo vision), between the stereo correspondences between two different time steps (visual odometry, map expansion), and between the stereo correspondences and the corresponding visible landmarks (SLAM).

12.4.2. Visual odometry

Visual odometry is used in the EKF robot motion model (see section 12.3.2) to estimate with high precision the relative pose of the robot between two time steps. The mean $\Delta\bar{\boldsymbol{\pi}}$ and the covariance matrix $\boldsymbol{P}_{\Delta_\pi}$ of the relative pose are calculated from the measurements $\boldsymbol{z}_{1,1}, \ldots, \boldsymbol{z}_{1,n}$ and $\boldsymbol{z}_{2,1}, \ldots, \boldsymbol{z}_{2,n}$ of two different times steps

$$\begin{aligned} f_{vo} : \mathbb{R}^{4\times n} \times \mathbb{R}^{4\times n} &\rightarrow \mathbb{R}^7 \times \mathbb{R}^{7\times 7}, \\ (\{\boldsymbol{z}_{1,1}, \ldots, \boldsymbol{z}_{1,n}\}, \{\boldsymbol{z}_{2,1}, \ldots, \boldsymbol{z}_{2,n}\}) &\mapsto (\Delta\boldsymbol{\pi}, \boldsymbol{P}_{\Delta_\pi}). \end{aligned} \tag{12.19}$$

Here $\boldsymbol{z}_{i,j} = (u_{i,j,l}, v_{i,j,l}, u_{i,j,r}, v_{i,j,r})^T$ denotes a stereo correspondence and $(\boldsymbol{z}_{1,j}, \boldsymbol{z}_{2,j})$ a correspondence pair between two time steps.

12.4.2.1. Outlier removal with RANSAC

Since the stereo correspondences or the correspondences between the time steps are not outlier free, RANSAC [138] is used. RANSAC is a repeating hypothesize-and-verify approach, which fits a model to the data to detect outliers.

In terms of our visual odometry, the model is represented by the relative robot pose $\Delta\boldsymbol{\pi}'$ and is instantiated from three corresponding stereo pairs

$$\begin{aligned} f_M : \mathbb{R}^{3\times 4} \times \mathbb{R}^{3\times 4} &\rightarrow \mathbb{R}^7, \\ ((\boldsymbol{z}_{1,j_1}, \boldsymbol{z}_{1,j_2}, \boldsymbol{z}_{1,j_3}), (\boldsymbol{z}_{2,j_1}, \boldsymbol{z}_{2,j_2}, \boldsymbol{z}_{2,j_3})) &\mapsto \Delta\boldsymbol{\pi}'. \end{aligned} \tag{12.20}$$

Each stereo pair is firstly reconstructed by our proposed refractive reconstruction function $\boldsymbol{p}_{i,j} = f_{rr}(\boldsymbol{z}_{i,j})$ (see section 6.2), and subsequently, the relative pose between the

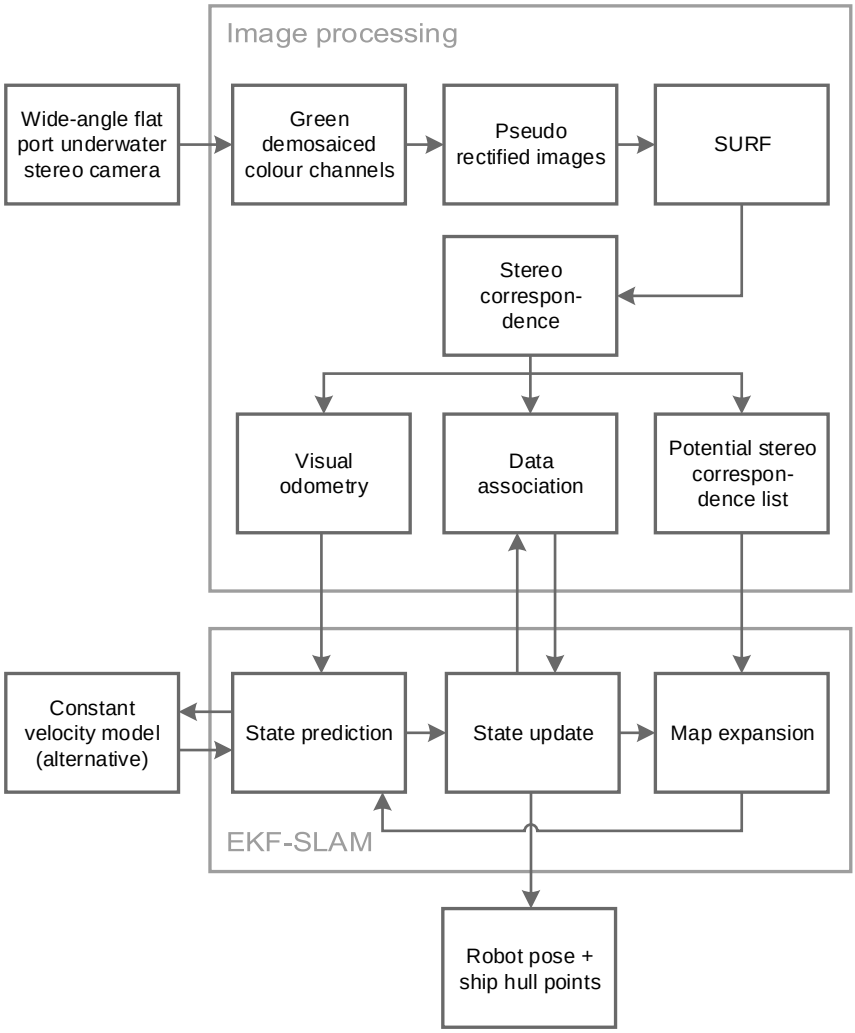

Figure 12.3.: The structure of the pose estimation algorithm

three reconstructed object points $(\boldsymbol{p}_{1,j_1}, \boldsymbol{p}_{1,j_2}, \boldsymbol{p}_{1,j_3})$ of time step 1 and the three reconstructed object points $(\boldsymbol{p}_{2,j_1}, \boldsymbol{p}_{2,j_2}, \boldsymbol{p}_{2,j_3})$ of time step 2 is computed by Horn's absolute orientation algorithm [115].

The fitness of a model is described by a distance function, which determines if a data point belongs to a model. The distance function in our visual odometry system is defined by

$$\begin{aligned} f_d : \mathbb{R}^3 \times \mathbb{R}^4 \times \mathbb{R}^7 &\rightarrow \{0,1\}, \\ (\boldsymbol{p}_{1,j}, \boldsymbol{z}_{2,j}, \Delta\boldsymbol{\pi}') &\mapsto b \end{aligned} \tag{12.21}$$

Here, based on the hypothesised relative robot pose $\Delta\boldsymbol{\pi}'$, the reconstructed object point $\boldsymbol{p}_1$ from time step 1 is projected onto the left and right camera images of time step 2 using the refractive forward projection function f_{rf} (see chapter 5)

$$\begin{aligned} \boldsymbol{u}'_{2,j,l} &= f_{rf}(\Delta\boldsymbol{\pi}' \oplus \boldsymbol{p}_{1,j}, l), \\ \boldsymbol{u}'_{2,j,r} &= f_{rf}(\Delta\boldsymbol{\pi}' \oplus \boldsymbol{p}_{1,j}, r). \end{aligned} \tag{12.22}$$

The projected image points $\boldsymbol{u}'_{2,l}$ and $\boldsymbol{u}'_{2,r}$ are finally compared with the actual observations $\boldsymbol{u}_{2,l}$ and $\boldsymbol{u}_{2,r}$ of time step 2

$$b = |\boldsymbol{u}'_{2,j,l} - \boldsymbol{u}_{2,j,l}| < c \wedge |\boldsymbol{u}'_{2,j,r} - \boldsymbol{u}_{2,j,r}| < c. \tag{12.23}$$

In our distance function, a correspondence pair of stereo correspondences is accepted, if the reprojection errors in both images are smaller than three times the standard deviation of the pixel noise, that is $c = 3\sigma_{uv}$.

12.4.2.2. Bundle adjustment

The high accuracy of the visual odometry is achieved by bundle adjustment. Based on the outlier free measurement sets from RANSAC, $\boldsymbol{z}'_{1,1}, \ldots, \boldsymbol{z}'_{1,m}$ and $\boldsymbol{z}'_{2,1}, \ldots, \boldsymbol{z}'_{1,m}$, bundle adjustment determines the optimal relative robot pose $\Delta\boldsymbol{\pi}$ and simultanously also the optimal position of the object points $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_m$ by minimising the total reprojection

error

$$\begin{aligned} \boldsymbol{x}^* &= \arg\min_{\boldsymbol{x}} e(\boldsymbol{x}, \boldsymbol{z}), \\ \boldsymbol{x} &= (\Delta\boldsymbol{\pi}, \boldsymbol{p}_1, \ldots, \boldsymbol{p}_m), \\ \boldsymbol{z} &= ((\boldsymbol{z}'_{1,1}, \ldots, \boldsymbol{z}'_{1,m}), (\boldsymbol{z}'_{2,1}, \ldots, \boldsymbol{z}'_{1,m})). \end{aligned} \tag{12.24}$$

The total reprojection error in the least-squares sense is defined as [116]

$$e(\boldsymbol{x}, \boldsymbol{z}) = \sum_{i=1}^{2} \sum_{j=1}^{n} \sum_{k \in \{l,r\}} \boldsymbol{e}_{i,j,k}^T \operatorname{diag}(\boldsymbol{\sigma}_{uv}^2)^{-1} \boldsymbol{e}_{i,j,k}, \tag{12.25}$$

and comprises the single reprojection errors $\boldsymbol{e}_{i,j,k}$

$$\boldsymbol{e}_{i,j,k} = \begin{cases} f_{rf}(\Delta\boldsymbol{\pi} \oplus \boldsymbol{p}_j, k) - \boldsymbol{u}_{i,j,k}, & \text{if } i = 1, \\ f_{rf}(\boldsymbol{p}_i, j) - \boldsymbol{u}_{i,j,k}, & \text{if } i = 2 \end{cases} \tag{12.26}$$

in all four camera images.

To enable visual odometry in real-time, we use the graph-based optimisation framework of Kümmerle et al. [117] and limit the number of pairs of stereo correspondences to $n \leq 500$. Additionally, the best RANSAC model with the corresponding reconstructed object points is used as initial guess. The covariance matrix $\boldsymbol{P}_{\Delta_\pi}$ is obtained from the information matrix in the graph-based estimation process.

12.4.3. Data association

In the context of SLAM, the data association relates the landmarks of the map to observations in the image. Particularly in EKF-SLAM, a correct data association is essential for a consistent estimation process. Otherwise, wrong data associations can lead to a chaotic behaviour and are non-reversible.

The first step of the data association has been described in section 12.4.1, where the correspondences between the visible landmarks ${}^w\boldsymbol{m}_1, \ldots, {}^w\boldsymbol{m}_n$ of the SLAM map and the observations $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ of the current time step were established by SURF. Similar to the correspondences for visual odometry, also these SLAM correspondences are contaminated with outliers, and thus are filtered by RANSAC.

The RANSAC model f_M and the distance function f_d are very similar to their counterparts in the visual odometry section (section 12.4.2). Rather than reconstructing two 3D point sets from the observations of the current and previous time step, now, the recon-

structed object points from the current time step based on the observations $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ and the landmarks ${}^w\boldsymbol{m}_1, \ldots, {}^w\boldsymbol{m}_n$ of the SLAM map are used for model instantiation

$$\begin{aligned} f_M : \mathbb{R}^{4\times 3} \times \mathbb{R}^{3\times 3} &\rightarrow \mathbb{R}^7, \\ ((\boldsymbol{z}_{i_1}, \boldsymbol{z}_{i_2}, \boldsymbol{z}_{i_3}), ({}^w\boldsymbol{m}_{i_1}, {}^w\boldsymbol{m}_{i_2}, {}^w\boldsymbol{m}_{i_3})) &\mapsto {}^r\boldsymbol{\pi}_w. \end{aligned} \tag{12.27}$$

The distance function is practically identical and only differs in the variables used

$$\begin{aligned} f_d : \mathbb{R}^3 \times \mathbb{R}^4 \times \mathbb{R}^7 &\rightarrow \{0, 1\}, \\ ({}^w\boldsymbol{m}, \boldsymbol{z}_i, {}^r\boldsymbol{\pi}_w) &\mapsto b. \end{aligned} \tag{12.28}$$

12.4.4. Potential stereo correspondence list

Since the number of landmarks in EKF-SLAM is limited by the quadratic complexity in the EKF update step, we keep the number of visible landmarks low and choose landmarks, which can be tracked over a longer period of time. To this end, only SURF features with a high response value are selected and the stereo correspondences are preselected using a potential stereo correspondence list. With this list, the stereo correspondences are tracked and evaluated for a certain time period to separate them from only sporadically visible correspondence pairs. To obtain a uniform distribution of the landmarks and thus a higher accuracy in the estimation process, new potential stereo correspondences need to keep a certain distance to visible landmarks and are randomly chosen from 4×3 different image sections.

12.5. Results

The pose estimation process was tested in a water tank to evaluate the performance of the proposed hull tracking, the pose estimation accuracy of the proposed refractive camera model-based underwater SLAM system, and the overall real-time performance under real conditions. To reproduce the ship hull surface in the tank, we went to the dry dock in Lyttelton Harbour [146] and filmed there the hull of a ship to be inspected. Some photos of the field work are depicted in Fig. 12.4. From the footage, we created a $0.84 \times 5.5\,\mathrm{m}^2$ large photomosaic, where a part of it is displayed in Fig. 12.5. The water tank including the underwater camera and the waterproof printout of the photomosaic are shown in Fig. 12.6.

To evaluate the accuracy of our pose estimation system, we attached the underwater

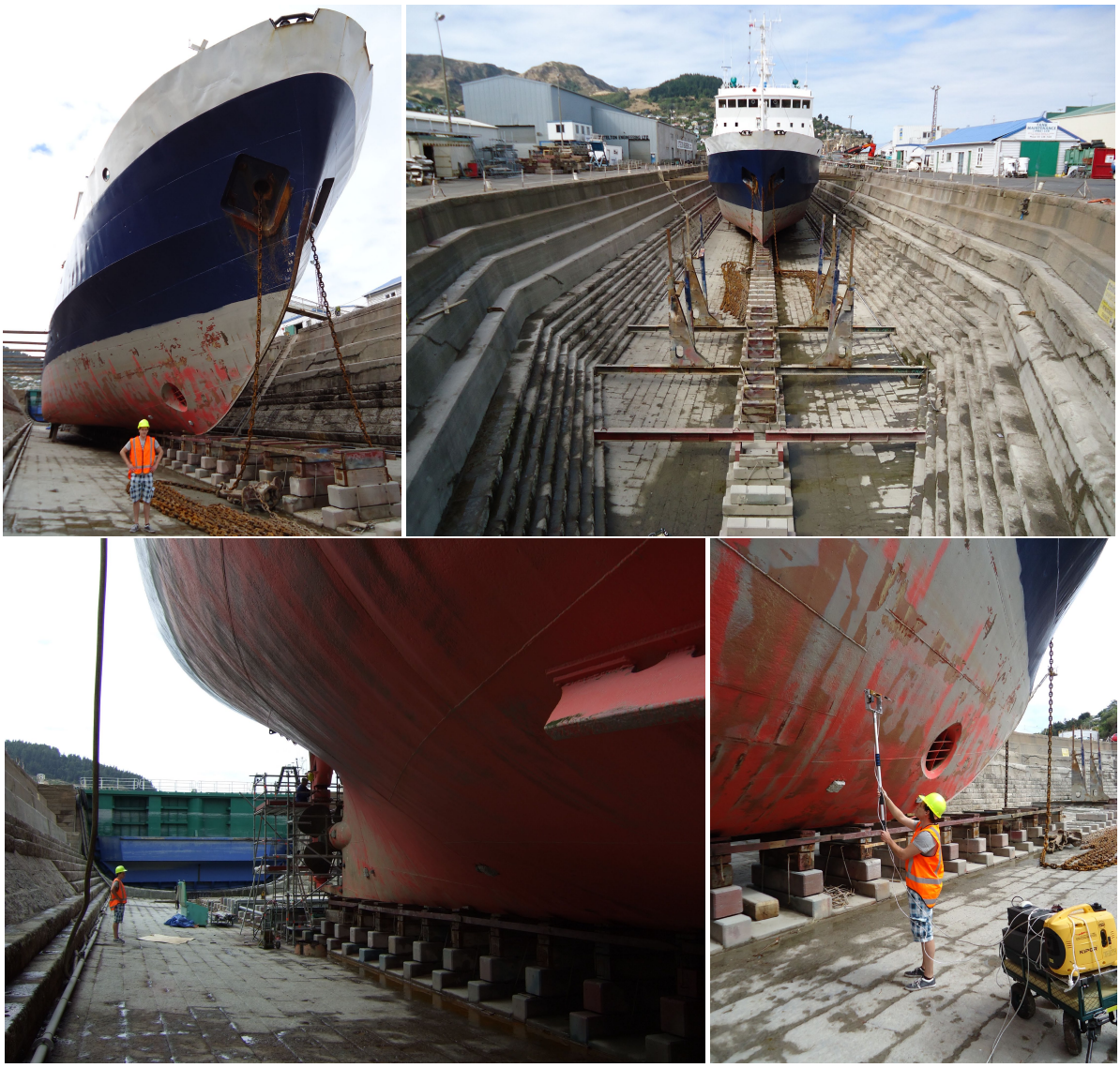

Figure 12.4.: Filming the hull of a ship in the dry dock of Lyttelton Harbour in Christchurch, New Zealand

Figure 12.5.: A part of the photomosaic, which was created from the dry dock footage: The figure shows that the ship hull surface is not completely smooth, but is sufficiently rough for visual tracking.

Figure 12.6.: Photos of the water tank experiment: The left photo shows the empty water tank with the $0.84 \times 5.5\,\mathrm{m}^2$ large ship hull printout and the underwater camera, which is attached to a movable carrier on top of the water tank. The ground truth was measured through the camera with a tape measure. The right photo shows the filled water tank and the partially submerge underwater camera while filming the printout. The distance of the camera to the printout is $\sim$65 cm.

camera to a carrier on top of the water tank and moved it in two loops over the photomosaic. The trajectory of the camera was constraint by the carrier to straight lines and limited by two limit switches. In this way, the ground truth of the camera pose could be obtained more easily than in experiments with a trajectory with more DOF. The displacement of the camera of 4.25 m was determined by filming a tape measure before the water tank was filled with water. The pool was illuminated by several fluorescent lamps.

For the observations, we assumed a pixel noise of $\sigma_{uv} = 1\,\text{px}$, and for the linear velocity robot motion model, which is very rarely used, a linear velocity noise of $\sigma_V = 5\,\text{cm/s}$ and a rotational velocity noise of $\sigma_\Omega = 0.01\,\text{rad/s}$.

To enable the robot pose estimation in real-time, we used the software libraries OpenCV (image processing, GPU SURF) [118], MRPT (EKF-SLAM) [147], g^{2}o (bundle adjustment) [117] and Eigen (linear algebra) [127]. We also used pipelining in which the calculation of the stereo correspondences was executed on a separate thread, and thus, the pose estimation was delayed by one frame. The algorithms were executed on an Intel Core i7-2600 CPU. Furthermore, the pose estimation algorithm was applied to a preprocessed video stream in which the green colour channel extraction, the demosaicing and the creation of pseudo rectified images had already been performed.

The experiment showed that the tracking of the ship hull surface is easily possible underwater under good visibility conditions. This is supported by Fig. 12.7 in which the number of stereo correspondences after the application of the epipolar constrained is shown. The number of stereo correspondences varies between approximately 1100 and 100 stereo correspondences. This demonstrates that also for regions with fewer SURF features, the number of stereo correspondences is still sufficiently high for the continuous tracking of the hull. To get an idea of the detected SURF stereo correspondences, an example stereo image is shown in Fig. 12.8 for the EKF landmarks tracked.

In Fig. 12.9 the estimated distance of the robot to its starting position including the 3-σ standard deviation is shown. The standard deviation σ_d was calculated by

$$\sigma_d = \sqrt{\frac{{}^w\boldsymbol{r}^T \boldsymbol{P}_{rr} {}^w\boldsymbol{r}}{{}^w\boldsymbol{r}^T {}^w\boldsymbol{r}}}, \tag{12.29}$$

where $\boldsymbol{P}_{rr}$ denotes the covariance matrix of the robot position. The figure demonstrates the high accuracy of the estimated distance. The error at the maximum distant point amounts to only 9 mm in the first loop and decreases to an error of only 7 mm after revisiting the point. This corresponds to a relative error of less than 2.1 ‰ or 1.7 ‰, re-

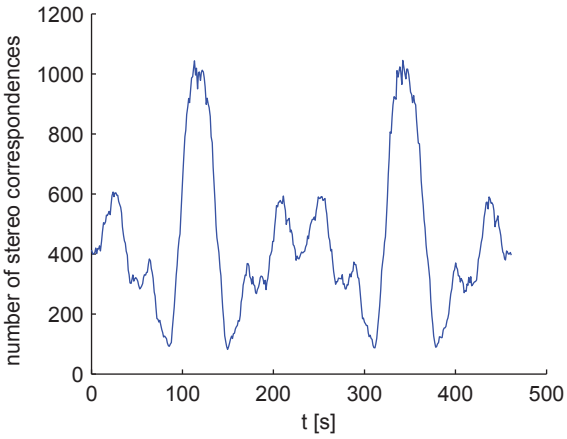

Figure 12.7.: The number of stereo correspondences, which were tracked during the camera motion over the ship hull printout: Here, the number of stereo correspondences after the application of the epipolar constraint is shown.

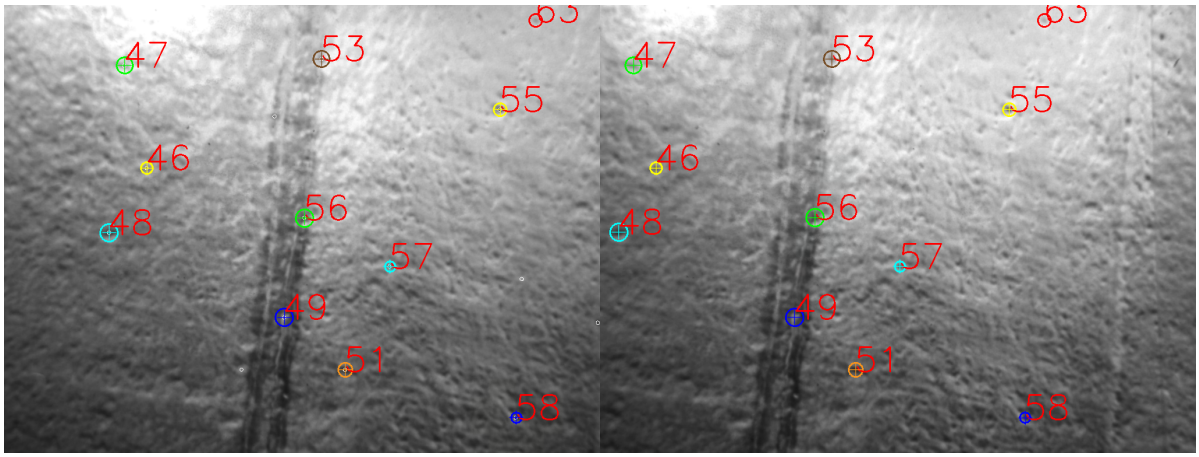

Figure 12.8.: EKF landmark tracking by SURF stereo correspondences: The coloured circles represent the SURF features and the numbers the corresponding landmark IDs. The 3σ-uncertainty ellipses (white) are also depicted. Due to the small uncertainty, which is mainly determined by the pixel noise, the size of the ellipses is very small.

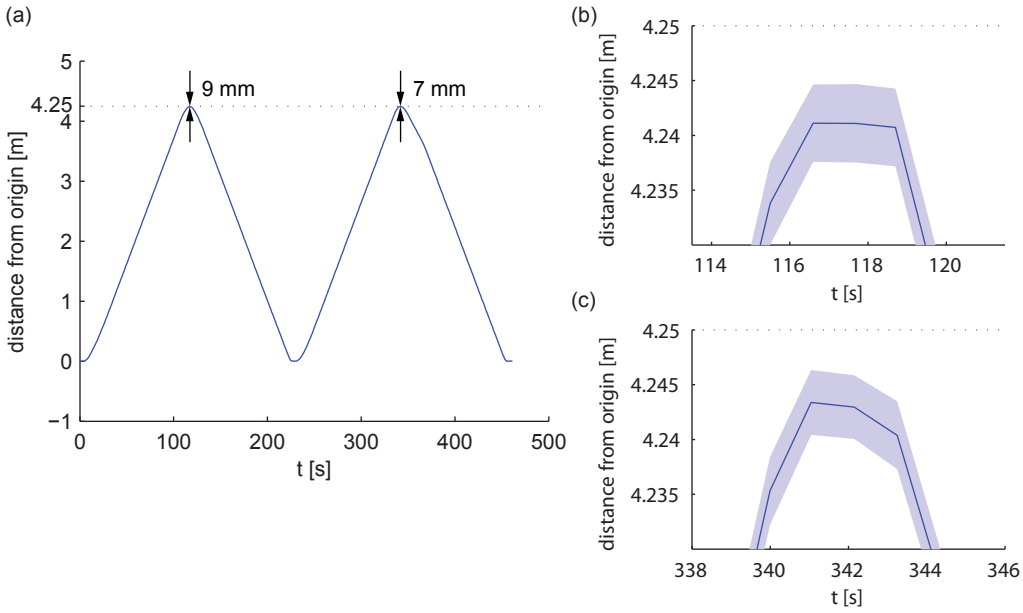

Figure 12.9.: (a) The estimated mean distance of the robot to its starting position with the 3-σ confidence interval, and the magnified views (b) and (c) of the maximum distant point at the two turning points of the loops. The estimated distance error is very small and corresponds to only 2.1 ‰ or 1.7 ‰.

spectively. According to the 3-σ confidence interval, the filter seems to be over-confident, as the ground truth of 4.25 m does not lie in the interval. However, it is also possible that the estimation process is more accurate than the ground truth measured.

The estimated orientation in the form of Euler angles with the corresponding 3-σ confidence intervals is shown in Fig. 12.10. For the conversion of the probability distribution of the rotation quaternion ${}^{w}\boldsymbol{q}_r$ to the probability distribution in Euler angels, we used [148]. The maximum error in the estimated mean orientation is for the yaw and pitch angle about 0.1° and for the roll angle 0.02°.The figures show a consistent estimation of the orientation. The ground truth of 0° stays always in the 3-σ confidence interval, which varies for the yaw angle between -0.53°-0.47°, for the pitch angle between -0.49°-0.52°, and for the yaw angle between -0.12°-0.12°. Consequently, the effects by an inaccurate estimation of the robot orientation in the EKF, as discussed in section 12.3.6, stay low.

The complete trajectory of the robot motion with the estimated mean and covariance of the robot pose at selected time steps, and the estimated final landmark map are depicted in Fig. 12.11. As expected from the constrained motion path, the estimated trajectory of the robot precisely follows a straight line. A deviation from the straight line motion is almost not recognisable and is only indicated by the slight offsets between

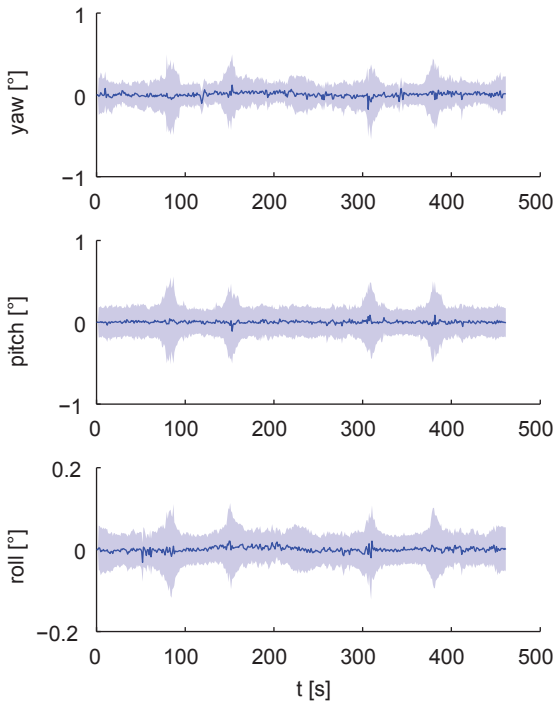

Figure 12.10.: The estimated mean yaw, pitch and roll angles during the motion with the 3-σ-confidence intervals.

the two loops of the trajectory. The distance between the turning points of the loops in the viewing direction of the camera (Fig. 12.10a) amounts to 1.8 cm, and is in the vertical direction (Fig. 12.10b) with only 4 mm significantly smaller. The magnified views (Fig. 12.10c and Fig. 12.10d) of the turning points also show the estimated uncertainties in the form of ellipses, which are not visible in Fig. 12.10a and Fig. 12.10b due to their small size. Here, the figures reveal an over-confidence of the estimates, which is larger in the viewing direction (Fig. 12.10c) of the camera. In a consistent estimation process, the uncertainty ellipses should overlap. There are numerous possible reasons for this inaccuracy. The small inaccuracy could be, for example, explained by the small distortion of the reconstructed object space, as demonstrated by our proposed evaluation of the refractive reconstruction accuracy (see chapter 8).

Fig. 12.12 shows the number of landmarks in the map with respect to time. The figure needs to be considered together with Fig. 12.13 in which the processing times during the robot motion are shown. The map size is directly related to the execution time of EKF update step, which represents, as discussed, the bottleneck in EKF-SLAM. The maximum maps size of 123 landmarks is approximately reached at the turning point of the first loop, where the robot starts revisiting the already explorer ship hull region.

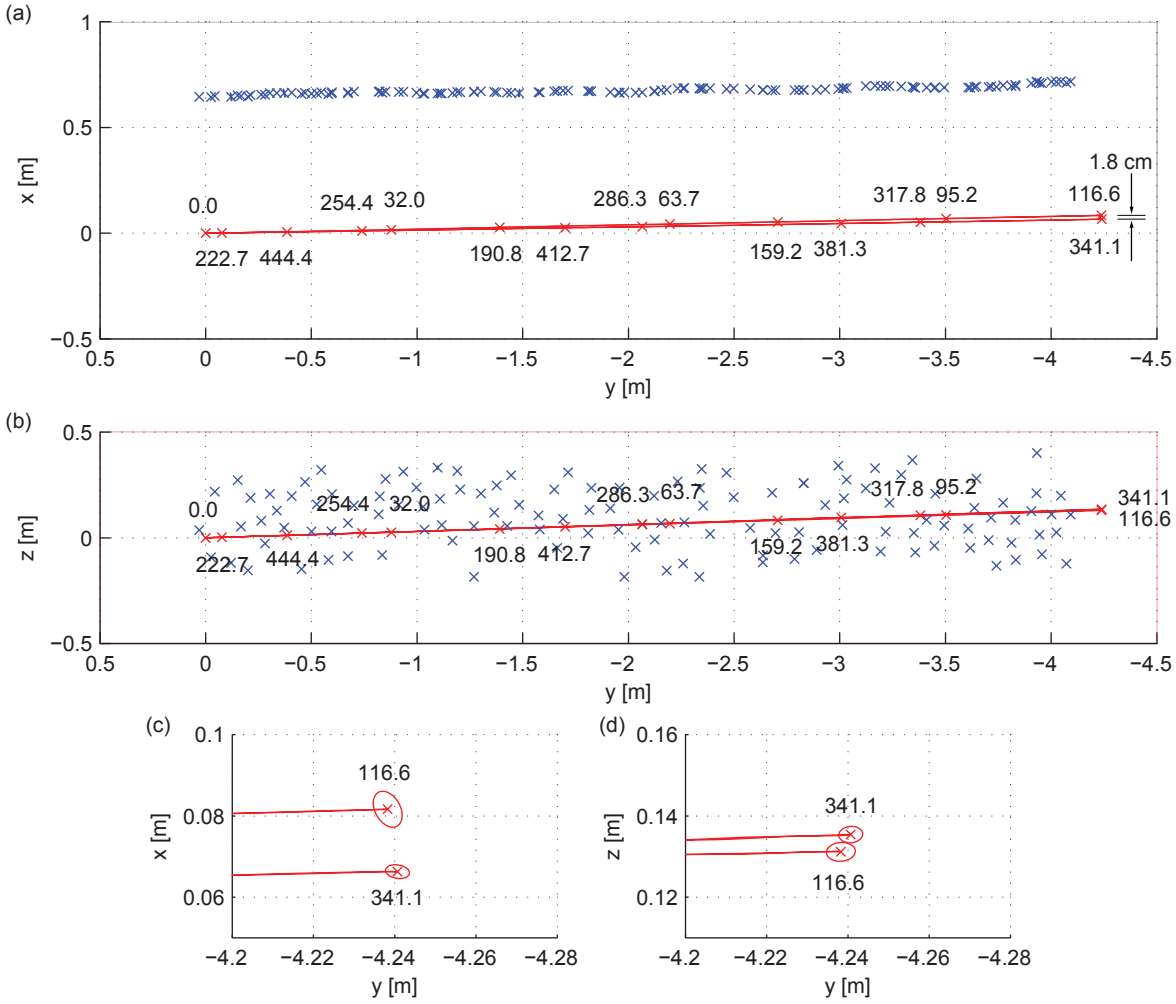

Figure 12.11.: 2D plots of the estimated robot trajectory (red line), the estimated robot position (red crosses with ellipses and timestamps in seconds) and the estimated final landmark positions (blue crosses with ellipses) (compare with the 3D plot of Fig. 12.2): (a) The orthographic view perpendicular to the optical axis and the baseline of the underwater camera. (b) The orthographic view in the viewing direction of the camera. (c) and (d) The magnified views of the most distant trajectory region to show the uncertainty regions in the form of ellipses.

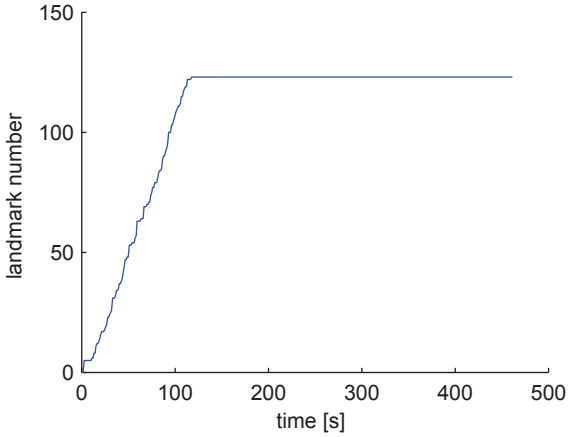

Figure 12.12.: The number of landmarks in the SLAM map

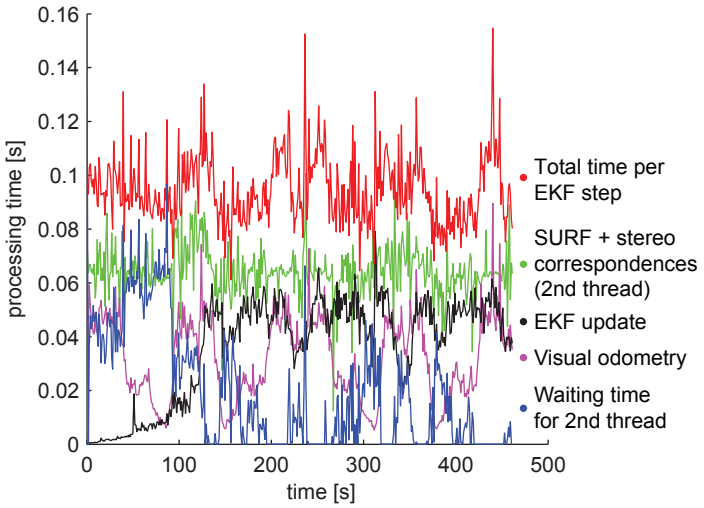

Figure 12.13.: The processing times of the pose estimation process

From this point, the processing time of the EKF update step stays on average constant and requires about 48 ms. The detection of SURF features including the generation of stereo correspondences represents with on average 64.5 ms the most time consuming processing step. For that reason, as discussed before, this processing step was moved to a separate thread, where the detection of SURF features is additionally executed on the GPU of our GTX 560 graphics card, which is equipped with 336 compute unified device architecture (CUDA) cores. We assume that the execution time of this step can be still significantly reduced with a more modern graphics card with more cores. The visual odometry represents the third largest time-consuming processing step. The execution time of the visual odometry is strongly correlated with the number of stereo correspondences (see Fig. 12.7). As mentioned before, in order to bound the execution time of the visual odometry, we have limited the number of stereo correspondences to 500, which results in an execution time of ∼50 ms. To show the load balancing between the main thread and the thread for the extraction of the stereo correspondences, we have also plotted the time, which the main thread has to wait until the second thread is finished. The figure shows that the waiting time is short when many stereo correspondences are available, and thus the processing of the visual odometry requires more time. This indicates that the visual odometry processing step could be also moved to another third thread. As shown in Fig. 12.13, the mean total execution time per EKF step is 95 ms or 10.5 Hz. If we assume this 3-step pipeline and the use of a more powerful graphics card, the EKF update step would represent the bottleneck, and the speed of our proposed pose estimation algorithm could be doubled to $1/48\,\text{Hz} = 21\,\text{Hz}$. This would enable the maximum frame rate of our camera of 20 Hz to be fully exploited.

12.6. Conclusion

In this chapter a novel accurate ship hull relative pose estimation system was proposed, which is based on the physics-based refractive camera model in combination with EKF-SLAM and bundle adjustment-based visual odometry. In a water tank experiment with good visibility conditions, this wide-angle underwater stereo vision system enabled the accurate pose estimation relative to a ship hull printout along a 4.25 m long straight-line trajectory. Here, the distance to the starting point was determined with a high precision of 2.1 ‰, and the orientation with a precision of 0.1°. Compared to the stereo vision pose estimation system of Negahdaripour and Firoozfam [24], our proposed ship hull pose estimation system is approximately one order of magnitude more accurate. Our

research also demonstrated that the continuous visual tracking of ship hulls underwater and in real-time is possible. In contrast to the work of Kim and Eustice [49] and Hover et al. [20], no sonar-based DVL is necessary under those conditions. Compared with their estimation framework, which runs with a frequency of 2-3 Hz, and despite the refractive camera model, our proposed pose estimation algorithm is several times faster and thus more suitable for the real-time navigation of autonomous underwater vehicles.

13. Conclusion

This thesis proposed novel methods for accurate high-resolution surface reconstruction and underwater localisation with a wide-angle flat port underwater stereo camera. To achieve these two main goals, a number of enabling methods were proposed.

Firstly, the effects, which occur due to the refraction of light at the flat port of an underwater housing, were described. Here, a novel fundamental refractive magnification function was proposed, which enables straight forward descriptions of 10 newly discovered and 9 known effects or methods. Furthermore, the large extent of chromatic aberration of about 10 px for a 1024×768 px wide-angle (HFOV=70°) underwater camera was revealed. This evaluation also showed that the thickness of the port needs to be modelled. Otherwise, a port thickness relative projection error of about 2 px/cm occurs.

Secondly, novel real-time, memory efficient, accurate refractive forward projection methods were successfully evaluated. The results of this research represent a significant breakthrough for real-time underwater vision. This thesis showed that a sufficiently accurate refractive forward projection of an object point is possible in only 105 ns using a lookup table-based approach. The method is two orders of magnitude faster than the often recommended 12th degree polynomial equation for thick ports. A more memory-efficient lookup table-based refractive forward projection method was also proposed and successfully evaluated. In this forward projection method, the lookup table size is independent of the maximum object distance and does not increase quadratically with it, as it is the case for the fastest 105 ns lookup table-based method. While being independent of the maximum object distance, the proposed lookup table-based method is only 11 % slower. For the case that lookup table-based methods are not suitable, such as refractive camera calibration or changing indices of refraction, this thesis also proposed an optimisation-based refractive forward projection method, which is about twice as fast as prior optimisation-based refractive forward projection methods.

Thirdly, the thesis evaluated the imaging and reconstruction accuracies under naturally varying indices of refraction. With respect to the ∼0.8 MP camera used, the water pressure had the strongest effect on projection and reconstruction (14.9 px, 18.3 mm at a

distance of 1 m), followed by the water salinity (11.3 px, 14 mm), the wavelength of light (9.8 px, 12.3 mm), the water temperature (2.3 px, 4.1 mm) and the air pressure in the underwater housing (0.4 px, 0.5 mm). These results clearly show that standard values for the index of refraction of water or air cannot be reliably used for accurate, high resolution, wide-angle flat port underwater stereo vision.

Fourthly, a multi-view reprojection error-based underwater camera calibration method was proposed. Depending on the number of stereo image pairs used, this calibration method is at least 3.5 to 5 times more accurate than prior refractive calibration methods. Due to the use of the proposed efficient and accurate refractive forward projection method, the calibration with 100 stereo image pairs was achieved in less than one second. The calibration was also tested under real-world conditions. Here, an inaccuracy of the physics-based underwater camera model was revealed, which, however, could be minimised by a proposed recalibration of the baseline of the wide-angle underwater stereo camera. As a result, the reprojection error was almost halved to 0.19 px.

Fifthly, the reconstruction accuracy of the physics-based underwater stereo camera with the proposed more accurate refractive indices, the more accurate refractive calibration and the corrected calibration of the stereo baseline was evaluated. In this evaluation, the 3D reconstruction at several distances was considered. Considering a range of 65-105 cm, a maximum mean reconstruction accuracy of 0.5 mm in lateral direction and of 1.8 mm in depth direction was measured in a viewing volume of $77 \times 55 \times 40\,\mathrm{cm}^3$ in front of the camera. For a flat $77 \times 55\,\mathrm{cm}^2$ front parallel plane at a distance of 65 cm, the reconstruction accuracy was about 0.2 mm in lateral and 0.6 mm in depth direction. For the reconstruction of smaller marine organisms on the hull of ships with a size of only few millimetres, this relative reconstruction error will be still smaller.

Sixthly, a more accurate single interface approximation was proposed. Here the calibration of the interface distance was based on the accurate two interface model and used two different interface distances for 2D imaging and 3D reconstruction. Applying this improved single interface approximation to the wide-angle underwater camera, the maximum imaging error could be reduced to about 0.9 px and the maximum reconstruction error to about 0.8 mm. However, these errors are still relatively high, and thus the use of the more accurate two interface physics-based underwater camera model is recommended.

Seventhly, two novel accurate, efficient methods to restore the surface images captured by a wide-angle flat port underwater stereo camera were presented. These methods enable the accurate real-time minimisation of chromatic aberration and the image

undistortion of captured object surfaces. These two techniques establish the foundation for highly accurate, high-resolution texture mapping of reconstructed underwater structures. The proposed algorithms demonstrated that the minimisation of chromatic aberration of an approximately 0.8 MP image can be achieved in less than 82 ms, and the correction of image distortion of an about 0.3 MP large image region in less than 15 ms.

Eighthly, two novel methods for facilitated (stereo) correspondence were presented. The first method proposed pseudo rectified images, which significantly reduce the strong image distortion in wide-angle underwater cameras. The second method represents an accurate and efficient approach to compute epipolar curves for constrained correspondence search. Using the two proposed methods in a real-world experiment, the mean and maximum measured epipolar constraint errors were about 0.19 px or 0.73 px, respectively. The calculation of 768 epipolar curves, each with 512 epipolar constraint errors, was performed in only 128 ms.

Finally, a novel method for accurate, real-time, ship hull relative pose estimation under simplified conditions was presented. This proposed pose estimation system is based on the physics-based refractive underwater camera model, accurate indices of refraction, bundle adjustment-based visual odometry and SLAM. The performance of the system was tested in a water tank with good visibility, a ship hull printout and a straight-line camera trajectory. Using a wide-angle flat port underwater stereo camera, the localisation system achieved highly accurate pose estimation accuracies with an error of only 2 ‰ in the estimated starting point distance and an error of only 0.1° in the estimated orientation. In contrast to prior research, this visual SLAM-based system uses the physics-based underwater camera model and proved efficient enough to exploit the full frame rate of the camera of 20 Hz, which enables the real-time navigation of autonomous underwater vehicles.

13.1. Future research

In this thesis, EKF-SLAM was used for the mapping of small ship hull regions. Due to the quadratic complexity in the number of landmarks, the map size of EKF-SLAM is limited. In order to keep the high processing frequency of the EKF and to enable larger maps, EKF-SLAM integrated as sub-maps into graph-based SLAM represents a possible solution. Furthermore, it is planned to create accurate, high-resolution 3D ship hull models using the SLAM maps and the surface image restoration methods proposed.

Another future goal of this research represents the automatic mapping and recognition of marine organism on the hulls of ships to prevent the introduction of invasive marine species into new biotopes.

A. Appendix

A.1. Detailed derivation of the refractive magnification function

In this section the detailed derivation of the refractive magnification function (4.7) on page 41 is shown. According to (4.1), the refractive magnification function is defined by

$$f_m(r') = \frac{r'}{r}.$$

For the derivation, let's start with (4.3):

$$r_p = r_{ap} + t\tan\theta_p + (z_p - t - d)\tan\theta_w.$$

Substituting r_p by (4.6), r_{ap} by (4.2) and using $\tan(x) = \frac{\sin x}{\sqrt{1-\sin^2 x}}$, we have:

$$\frac{rz_p}{f} = \frac{r'd}{f} + t\frac{\sin\theta_p}{\sqrt{1-\sin^2\theta_p}} + (z_p - t - d)\frac{\sin\theta_w}{\sqrt{1-\sin^2\theta_w}}. \tag{A.1}$$

From Snell's law (4.4) follows $\sin\theta_i = \frac{n_a}{n_i}\sin\theta_a$ with $i \in \{p, w\}$. Substituting $\sin\theta_a$ by (4.5) yields $\sin\theta_i = \frac{n_a}{n_i}\frac{r'}{\sqrt{r'^2+f^2}}$. Substituting this term into (A.1) leads to:

$$\frac{rz_p}{f} = \frac{r'd}{f} + t\frac{\frac{n_a}{n_p}\frac{r'}{\sqrt{r'^2+f^2}}}{\sqrt{1-\frac{n_a^2}{n_p^2}\frac{r'^2}{r'^2+f^2}}} + (z_p - t - d)\ldots. \tag{A.2}$$

Let's denote the highlighted term c_p and simplify it (common denominator, compound fraction):

$$c_p = \frac{n_a}{n_p}\sqrt{\frac{n_p^2(r'^2+f^2)}{(r'^2+f^2)(n_p^2(r'^2+f^2) - n_a^2 r'^2)}}. \tag{A.3}$$

A further simplification by a reduction of the fraction (see highlighted terms) leads to

$$c_p = \frac{n_a}{\sqrt{n_p^2 f^2 + (n_p^2 - n_a^2) r'^2}}. \tag{A.4}$$

Substituting (A.4) into (A.2) and defining c_w similar to c_p yields:

$$\frac{r z_p}{f} = \frac{r' d}{f} + t r' c_p + (z_p - t - d) r' c_w. \tag{A.5}$$

Transforming the equation to $\frac{r'}{r}$ according to (4.1) results in

$$\frac{r'}{r} = \frac{z_p}{d + t f c_p + (z_p - t - d) f c_w}. \tag{A.6}$$

Using $b_i = f c_i$ with $i \in \{p, w\}$ and rearranging the denominator results in the refractive magnification function (4.7).

A.2. The pose compounding operator $\oplus$ and the inverse pose operator $\ominus$

The pose compounding operator $\oplus$ and the inverse pose operator $\ominus$ are used for transforming translations/points, rotations/orientations or poses/coordinate systems between coordinate systems in a brief convenient form.

A pose $\boldsymbol{\pi}$ is defined by its position and orientation. In its most general form, the pose compounding operator $\oplus$ transforms a pose ${}^b\boldsymbol{\pi}_c$ given in the coordinate system b to the coordinate system a using the relative pose ${}^a\boldsymbol{\pi}_b$

$${}^a\boldsymbol{\pi}_c = {}^a\boldsymbol{\pi}_b \oplus {}^b\boldsymbol{\pi}_c, \tag{A.7}$$

where ${}^a\boldsymbol{\pi}_b$ denotes the pose of b with respect to a.

The inverse pose operator $\ominus$ maps a pose ${}^b\boldsymbol{\pi}_a$ to its inverse ${}^a\boldsymbol{\pi}_b$

$${}^a\boldsymbol{\pi}_b = \ominus {}^b\boldsymbol{\pi}_a. \tag{A.8}$$

There is also a binary operator $\ominus$, which is similar to the binary operator $\oplus$, but uses the inverse pose of the second argument. It is defined as a combination of the binary $\oplus$

operator and the unary ⊖ operator

$$ {}^{a}\boldsymbol{\pi}_c = {}^{a}\boldsymbol{\pi}_b \ominus {}^{c}\boldsymbol{\pi}_b = {}^{a}\boldsymbol{\pi}_b \oplus (\ominus {}^{c}\boldsymbol{\pi}_b). \tag{A.9} $$

The arguments of ⊕ and ⊖ can also be represented by only a translation vector or an orientation instead of a full pose. This makes sense, if the transformation between two coordinate systems is only a translation or rotation, or if the object to be transformed is only a position/point/translation or orientation/rotation.

In addition, the orientation of a pose can have arbitrary representations, such as in the form of a rotation matrix, Euler angles, an angle axis representation or a rotation quaternion.

The definition of ⊕ and ⊖ for the 3-DOF case (one rotation, two translations) is found in [149], which is often cited in the context of six DOF. Definitions for six DOF of very similar operators can be found in the MRPT library [147, 148] (see Pose3d; ⊖ used as binary operator) and the Eigen [127] library (see Eigen::Transform; ⊖ by explicit inverse of the pose).

In the following, an example of the definition of ⊕ and ⊖ with translation vectors and rotation matrices is given. For ⊕, we have

$$ ({}^{a}\boldsymbol{t}_c, {}^{a}\boldsymbol{R}_c) = ({}^{a}\boldsymbol{t}_b, {}^{a}\boldsymbol{R}_b) \oplus ({}^{b}\boldsymbol{t}_c, {}^{b}\boldsymbol{R}_c), \tag{A.10} $$

where

$$ {}^{a}\boldsymbol{R}_c = {}^{a}\boldsymbol{R}_b \, {}^{b}\boldsymbol{R}_c, \tag{A.11} $$

$$ {}^{a}\boldsymbol{t}_c = {}^{a}\boldsymbol{R}_b \, {}^{b}\boldsymbol{t}_c + {}^{a}\boldsymbol{t}_b. \tag{A.12} $$

And ⊖ is defined as follows:

$$ ({}^{a}\boldsymbol{t}_b, {}^{a}\boldsymbol{R}_b) = \ominus({}^{b}\boldsymbol{t}_a, {}^{b}\boldsymbol{R}_a), \tag{A.13} $$

where

$$ {}^{a}\boldsymbol{R}_b = {}^{b}\boldsymbol{R}_a^T, \tag{A.14} $$

$$ {}^{a}\boldsymbol{t}_b = -{}^{a}\boldsymbol{R}_b \, {}^{b}\boldsymbol{t}_a. \tag{A.15} $$

Bibliography

[1] H.-G. Maas, "New developments in multimedia photogrammetry," *Optical 3D measurement techniques III*, 1995.

[2] R. Kawahara, S. Nobuhara, and T. Matsuyama, "A pixel-wise varifocal camera model for efficient forward projection and linear extrinsic calibration of underwater cameras with flat housings," in *Proc. of ICCV2013 Workshop on Underwater Vision*, 2013, pp. 819–824.

[3] MAF Biosecurity. (2011, June) The MAF Biosecurity New Zealand. [Online]. Available: www.biosecurity.govt.nz

[4] R. Piola and C. Conwell, "Vessel biofouling as a vector for the introduction of non-indigenous marine species to new zealand: Fishing vessels," MAF Biosecurity New Zealand, Tech. Rep., 2010.

[5] A. Bell, S. Phillips, E. Georgiades, and D. Kluza, "Risk analysis: Vessel biofouling," MAF Biosecurity NZ, Tech. Rep., 2011.

[6] IMO, *Guidelines for the control and management of ships' biofouling to minimize the transfer of invasive aquatic species*, The International Maritime Organization Std., 2011.

[7] A. Coutts and T. Dodgshun, "The nature and extent of organisms in vessel seachests: A protected mechanism for marine bioinvasions," *Marine pollution bulletin*, vol. 54, no. 7, pp. 875–886, 2007.

[8] L. T. Takata, M. B. Falkner, and S. Gilmore, "Commercial vessel fouling in california," California State Lands Commission, Tech. Rep., 2006.

[9] S. Raaymakers, "The ballast water problem: Global ecological, economic and human health impacts paper," International Maritime Organization, Tech. Rep., 2002, presented at the RECSO/IMO Joint Seminar on Tanker Ballast Water Management & Technologies Dubai, UAE 16-18.

[10] S. Gollasch, "The importance of ship hull fouling as a vector of species introductions into the North Sea," *Biofouling*, 2002.

[11] O. Floerl and A. Coutts, "Feasibility of using remote-operated vehicles (rovs) for vessel biofouling inspections," Department of Fisheries, Western Australia, Tech. Rep., 2013.

[12] J. Kinsey, R. Eustice, and L. Whitcomb, "A survey of underwater vehicle navigation: Recent advances and new challenges," in *Proceedings of the 7th Conference on Maneuvering and Control of Marine Craft (MCMC'2006). IFAC, Lisbon*, 2006.

[13] D. Ribas, P. Ridao, J. Tardós, and J. Neira, "Underwater slam in man-made structured environments," *Journal of Field Robotics*, vol. 25, no. 11-12, pp. 898–921, 2008.

[14] N. P. Fofonoff and R. C. Millard, "Algorithms for computation of fundamental properties of seawater," Unesco, Tech. Rep., 1983.

[15] O. J. Woodman, "An introduction to inertial navigation," University of Cambridge, Tech. Rep., 2007.

[16] A. V. Inzartsev, Ed., *Underwater Vehicles*. InTech, 2009.

[17] L. Chen, S. Wang, K. McDonald-Maier, and H. Hu, "Towards autonomous localization and mapping of auvs - a survey," *International Journal of Intelligent Unmanned Systems*, vol. 1, pp. 97–120, 2013.

[18] Teledyne RD Instruments. (2014, July) Workhorse navigator doppler velocity log (dvl) - precision navigation for the marine environment. [Online]. Available: http://www.rdinstruments.com/pdfs/wh_navigator_ds_lr.pdf

[19] E. Belcher, W. Hanot, and J. Burch, "Dual-frequency identification sonar (DIDSON)," in *Underwater Technology, 2002. Proceedings of the 2002 International Symposium on*. IEEE, 2002, pp. 187–192.

[20] F. S. Hover, R. M. Eustice, A. Kim, B. Englot, H. Johannsson, M. Kaess, and J. J. Leonard, "Advanced perception, navigation and planning for autonomous in-water ship hull inspection," *International Journal of Robotics Research, Special Issue on 3D Exploration, Mapping, and Surveillance*, vol. 31, no. 12, pp. 1445–1464, Oct 2012.

[21] soundmetrics. (2014, July) Didson sonars. [Online]. Available: http://www.soundmetrics.com/Products/DIDSON-Sonars

[22] M. VanMiddlesworth, M. Kaess, F. Hover, and J. Leonard, "Mapping 3d underwater environments with smoothed submaps," in *Conf. on Field and Service Robotics (FSR)*, Brisbane, Australia, Dec 2013.

[23] C. Mobley, *Light and water: radiative transfer in natural waters*. Academic Press, 1994.

[24] S. Negahdaripour and P. Firoozfam, "An ROV stereovision system for ship-hull inspection," *IEEE Journal of Oceanic Engineering*, vol. 31, no. 3, p. 551, 2006.

[25] Y. Swirski, Y. Y. Schechner, B. Herzberg, and S. Negahdaripour, "Caustereo: Range from light in nature," *Appl. Opt.*, vol. 50, no. 28, pp. F89–F101, Oct 2011.

[26] Y. Swirski and Y. Y. Schechner, "Caustereo: structure from underwater flickering illumination," in *Proc. SPIE*, vol. 8480, 2012, pp. 84 800D–84 800D–8.

[27] Z. Abidin and M. Arshad, "Visual Servoing with application to ROV for Ship Hull Inspection," in *Proceedings of International Conference on Man-Machine Systems*, 2006.

[28] P. Drews, J. Longui, and V. Rosa, "Real-time depth estimation for underwater inspection using dual laser and camera," in *Computing and Automation for Offshore Shipbuilding (NAVCOMP), 2013 Symposium on*, March 2013, pp. 52–56.

[29] F. Bruno, G. Bianco, M. Muzzupappa, S. Barone, and A. Razionale, "Experimentation of structured light and stereo vision for underwater 3d reconstruction," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 4, pp. 508–518, 2011.

[30] S. Harris, E. Slate, I. Inc, and M. Cataumet, "Lamp Ray: ship hull assessment for value, safety and readiness," *Oceans'99 MTS/IEEE. Riding the Crest into the 21st Century*, vol. 1, 1999.

[31] G. Inglis, C. Smart, I. Vaughn, and C. Roman, "A pipeline for structured light bathymetric mapping," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, Oct 2012, pp. 4425–4432.

[32] E. Ozsvald. (2014, June) kinect + water. [Online]. Available: http://www.fiveyearstobefamous.com/eszterOzsvald/kinect-water/

[33] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part I," *IEEE Robotics & Automation Magazine*, vol. 13, no. 2, pp. 99–110, 2006.

[34] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (SLAM): part II," *Robotics & Automation Magazine, IEEE*, vol. 13, no. 3, pp. 108–117, 2006.

[35] SeaBotix. (2010) Lbv300-5 crawler.

[36] G. Trimble and E. Belcher, "Ship berthing and hull inspection using the cetusii auv and miris high-resolution sonar," in *OCEANS'02 MTS/IEEE*, vol. 2. IEEE, 2002, pp. 1172–1175.

[37] Desert Star Systems. (2002) Ship hull inspections with aquamap. [Online]. Available: http://www.desertstar.com/

[38] J. Vaganay, M. Elkins, S. Willcox, F. Hover, R. Damus, S. Desset, J. Morash, and V. Polidoro, "Ship hull inspection by hull-relative navigation and control," in *OCEANS, 2005. Proceedings of MTS/IEEE.* IEEE, 2005, pp. 761–766.

[39] J. Vaganay, M. Elkins, D. Esposito, W. O'Halloran, F. Hover, and M. Kokko, "Ship hull inspection with the HAUV: US Navy and NATO demonstrations results," in *OCEANS 2006.* IEEE, 2006, pp. 1–6.

[40] S. Reed, A. Cormack, K. Hamilton, I. Ruiz, and D. Lane, "Automatic Ship Hull Inspection using Unmanned Underwater Vehicles," in *Proceedings from the 7th International Symposium on Technology and the Mine Problem. Monterey, USA*, 2006.

[41] M. Walter, F. Hover, and J. Leonard, "Slam for ship hull inspection using exactly sparse extended information filters," in *IEEE International Conference on Robotics and Automation (ICRA).* IEEE, 2008, pp. 1463–1470.

[42] J. Vaganay, L. Gurfinkel, M. Elkins, D. Jankins, and K. Shurn, "Hovering autonomous underwater vehicle-system design improvements and performance evaluation results," in *Proceedings of the International Symposium on Unmanned Untethered Submersible Technology (UUST)*, 2009, pp. 1–14.

[43] B. Englot, H. Johannsson, and F. Hover, "Perception, stability analysis, and motion planning for autonmous ship hull inspection," in *Proceedings of the International Symposium on Unmanned Untethered Submersible Technology (UUST)*, 2009.

[44] G. Hollinger, B. Englot, F. Hover, U. Mitra, and G. Sukhatme, "Uncertainty-driven view planning for underwater inspection," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, May 2012, pp. 4884–4891.

[45] H. Johannsson, "Toward lifelong visual localization and mapping," Ph.D. dissertation, Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, 2013.

[46] Desert Star Systems. (2014, July) Aquamap shiphull inspection. [Online]. Available: http://www.desertstar.com/aquamap.html

[47] A. Kim and R. Eustice, "Real-time visual slam for autonomous underwater hull inspection using visual saliency," *IEEE Transactions on Robotics*, vol. 29, no. 3, pp. 719–733, 2013.

[48] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. Leonard, and F. Dellaert, "isam2: Incremental smoothing and mapping with fluid relinearization and incremental variable reordering," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, may 2011, pp. 3281 –3288.

[49] A. Kim and R. Eustice, "Pose-graph visual SLAM with geometric model selection for autonomous underwater ship hull inspection," in *Proceedings of the 2009 IEEE/RSJ international conference on Intelligent robots and systems*. IEEE Press, 2009, pp. 1559–1565.

[50] P. Ridao, M. Carreras, D. Ribas, and R. Garcia, "Visual inspection of hydroelectric dams using an autonomous underwater vehicle," *Journal of Field Robotics*, vol. 27, no. 6, pp. 759–778, 2010.

[51] A. Hogue, A. German, and M. Jenkin, "Underwater environment reconstruction using stereo and inertial data," in *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on*, oct. 2007, pp. 2372 –2377.

[52] P. Grey, "Bumblebee2 datasheet," June 2012. [Online]. Available: http://www.ptgrey.com/support/downloads/10132

[53] Sony, "ICX204AK datasheet," Sony.

[54] Canon. (2014, November) Fixed focal length - ef lenses. Canon. [Online]. Available: http://www.canon.co.uk/For_Home/Product_Finder/Cameras/EF_Lenses/Fixed_Focal_Length

[55] Nikon. (2014, November) Nikkor camera lenses. Nikon. [Online]. Available: http://www.nikonusa.com/en/Nikon-Products/Camera-Lenses/All-Lenses

[56] H. Karara and L. Adams, *Non-Topographic Photogrammetry*, ser. Science and Engineering Series. American Society for Photogrammetry and Remote Sensing, 1989.

[57] M. Edge, *The underwater photographer*. Amsterdam Boston: Focal, 2010.

[58] Z. Wakimoto, "On designing underwater camera lenses," *Photogrammetric Engineering*, vol. 33(8), pp. 925–936, 1967.

[59] C. Kunz and H. Singh, "Hemispherical refraction and camera calibration in underwater vision," in *OCEANS 2008*, sept. 2008, pp. 1 –7.

[60] D. C. Brown, "Decentering distortion of lenses," *Photometric Engineering*, vol. 32, no. 3, pp. 444–462, 1966.

[61] G. McNeil, *Optical Fundamentals of Underwater Photography*. Photogrammetry, 1968.

[62] A. Ivanoff, Y. Le Grand, and P. Cuvier, "Optical system for distortionless underwater vision," U.S. Patent 2 730 014, January, 1956.

[63] A. Ivanoff and P. Cherney, "Correcting lenses for underwater use," *Journal of the SMPTE*, vol. 69, no. 4, pp. 264–266, 1960. [Online]. Available: http://journal.smpte.org/content/69/4/264.abstract

[64] J. Fryer and C. Fraser, "On the calibration of underwater cameras," *The Photogrammetric Record*, vol. 12, no. 67, pp. 73–85, 1986.

[65] Nikon. (2013, November) Evolution of nikonos. [Online]. Available: http://imaging.nikon.com/history/chronicle/history-nikonos/

[66] Y.-H. Kwon and J. B. Casebolt, "Effects of light refraction on the accuracy of camera calibration and reconstruction in underwater motion analysis," *Sports biomechanics*, vol. 5, no. 2, pp. 315–340, 2006.

[67] Y.-H. Kwon and S. Lindley, "Applicability of four localized-calibration methods in underwater motion analysis," in *XVIII International Symposium on Biomechanics in Sports*, 2000.

[68] T. Treibitz, Y. Schechner, and H. Singh, "Flat refractive geometry," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, june 2008, pp. 1 –8.

[69] V. Chari and P. Sturm, "Multiple-View Geometry of the Refractive Plane," in *20th British Machine Vision Conference (BMVC '09)*. London, Royaume-Uni: The British Machine Vision Association (BMVA), 2009.

[70] J. Gedge, M. Gong, and Y. Yang, "Refractive epipolar geometry for underwater stereo matching," in *Computer and Robot Vision (CRV), 2011 Canadian Conference on*. IEEE, 2011, pp. 146–152.

[71] T. Treibitz, Y. Schechner, C. Kunz, and H. Singh, "Flat refractive geometry," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 1, pp. 51–65, 2012.

[72] T. Yau, M. Gong, and Y.-H. Yang, "Underwater camera calibration using wavelength triangulation," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, 2013, pp. 2499–2506.

[73] A. Agrawal, S. Ramalingam, Y. Taguchi, and V. Chari, "A theory of multi-layer flat refractive geometry," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, june 2012, pp. 3346 –3353.

[74] R. E. Hopkins and H. E. Edgerton, "Lenses for underwater photography," *Deep Sea Research (1953)*, vol. 8, no. 3–4, pp. 312 – 317, 1961.

[75] A. Sedlazeck and R. Koch, "Calibration of housing parameters for underwater stereo-camera rigs," in *Proceedings of the British Machine Vision Conference*. BMVA Press, 2011, pp. 118.1–118.11.

[76] E. Hecht, *Optics (4th Edition)*, 4th ed. Addison Wesley, August 2001.

[77] A. Jordt-Sedlazeck, D. Jung, and R. Koch, "Refractive plane sweep for underwater images," in *Pattern Recognition*, ser. Lecture Notes in Computer Science, J. Weickert, M. Hein, and B. Schiele, Eds. Springer Berlin Heidelberg, 2013, vol. 8142, pp. 333–342.

[78] A. Jordt, "Underwater 3d reconstruction based on physical models for refraction and underwater light propagation," Ph.D. dissertation, Kiel University, 2014.

[79] A. Jordt-Sedlazeck and R. Koch, "Refractive structure-from-motion on underwater images," in *Computer Vision (ICCV), 2013 IEEE International Conference on*, Dec 2013, pp. 57–64.

[80] ——, "Refractive calibration of underwater cameras," in *ECCV (5)*, 2012, pp. 846–859.

[81] L. Kang, L. Wu, and Y.-H. Yang, "Experimental study of the influence of refraction on underwater three-dimensional reconstruction using the svp camera model," *Appl. Opt.*, vol. 51, no. 31, pp. 7591–7603, Nov 2012.

[82] ——, "Two-view underwater structure and motion for cameras under flat refractive interfaces," in *Computer Vision – ECCV 2012*, ser. Lecture Notes in Computer Science, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds. Springer Berlin Heidelberg, 2012, vol. 7575, pp. 303–316.

[83] G. Glaeser and H. Schröcker, "Reflections on refractions," *Journal for Geometry and Graphics*, vol. 4, no. 1, pp. 1–18, 2000.

[84] R. Kotowski, "Phototriangulation in multi-media photogrammetry," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1988. [Online]. Available: http://www.isprs.org/proceedings/XXVII/congress/part5/

[85] D. Herbison-Evans, "Solving quartics and cubics for graphics," in *Graphics Gems V*, A. W. Paeth, Ed. Academic Press, 1995, pp. 3–15.

[86] M. Abramowitz and I. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Dover publications, 1964.

[87] Intel, *Intel 64 and IA-32 Architectures Optimization Reference Manual*, Intel, April 2012.

[88] (2014, January) Quake III Arena. [Online]. Available: https://github.com/id-Software/Quake-III-Arena

[89] M. Robertson, "A brief history of invsqrt," Undergraduate Honors Thesis, University of New Brunswick, 2012.

[90] I. Volobouev. (2014, February) Fftjet. [Online]. Available: http://fftjet.hepforge.org/svn/trunk/fftjet/quartic_lib.cc

[91] J. G. F. Francis, "The qr transformation a unitary analogue to the lr transformation—part 1," *The Computer Journal*, vol. 4, no. 3, pp. 265–271, 1961.

[92] M. Jenkins and J. Traub, "A three-stage algorithm for real polynomials using quadratic iteration," *SIAM Journal on Numerical Analysis*, vol. 7, no. 4, pp. 545–566, 1970.

[93] L. Bartholdi. (2007, July) Translation of the jenkins-traub algorithm to c. [Online]. Available: http://www.uni-math.gwdg.de/laurent/FR/fr/src/rpoly.c

[94] GSL. (2011, May) GSL - GNU Scientific Library 1.15. [Online]. Available: http://www.gnu.org/software/gsl/

[95] R. Li, H. Li, W. Zou, R. Smith, and T. Curran, "Quantitative photogrammetric analysis of digital underwater video imagery," *Oceanic Engineering, IEEE Journal of*, vol. 22, no. 2, pp. 364–375, 1997.

[96] Y. Schechner and N. Karpel, "Clear underwater vision," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 1, june-2 july 2004, pp. I–536 – I–543 Vol.1.

[97] Y.-J. Chang and T. Chen, "Multi-view 3d reconstruction for scenes under the refractive plane with known vertical direction," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 351–358.

[98] G. Telem and S. Filin, "Photogrammetric modeling of underwater environments," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 5, pp. 433 – 444, 2010.

[99] G. T. McNeil, "Metrical fundamentals of underwater lens system," *Optical Engineering*, vol. 16, no. 2, pp. 162 128–162 128, 1977.

[100] A. Sedlazeck and R. Koch, "Perspective and non-perspective camera models in underwater imaging — overview and error analysis," in *Proceedings of the 15th international conference on Theoretical Foundations of Computer Vision: outdoor and large-scale real-world scene analysis*. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 212–242.

[101] M. R. Shortis and E. S. Harvey, "Design and calibration of an underwater stereo-video system for the monitoring of marine fauna populations," *International Archives of Photogrammetry and Remote Sensing*, vol. 32, pp. 792–799, 1998.

[102] M. P. M. Reddy, *Descriptive Physical Oceanography*. Oxford and IBH Publishing Company, 2001.

[103] L. Talley, G. Pickard, W. Emery, and J. Swift, *Descriptive Physical Oceanography: An Introduction*. Elsevier Science, 2011.

[104] B. Boehrer and M. Schultze, "Stratification of lakes," *Reviews of Geophysics*, vol. 46, no. 2, 2008.

[105] S. N. Kasarova, N. G. Sultanova, C. D. Ivanov, and I. D. Nikolov, "Analysis of the dispersion of optical plastic materials," *Optical Materials*, vol. 29, no. 11, pp. 1481 – 1490, 2007.

[106] J. M. Cariou, J. Dugas, L. Martin, and P. Michel, "Refractive-index variations with temperature of pmma and polycarbonate," *Appl. Opt.*, vol. 25, no. 3, pp. 334–336, Feb 1986.

[107] P. E. Ciddor, "Refractive index of air: new equations for the visible and near infrared," *Appl. Opt.*, vol. 35, no. 9, pp. 1566–1573, Mar 1996.

[108] COESA, *US standard atmosphere, 1976*. NOAA, NASA, and USAF, 1976.

[109] I. Mahon, O. Pizarro, M. Johnson-Roberson, A. Friedman, S. Williams, and J. Henderson, "Reconstructing pavlopetri: mapping the world's oldest submerged town using stereo-vision," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2011.

[110] M. Johnson-Roberson, O. Pizarro, S. Williams, and I. Mahon, "Generation and visualization of large-scale three-dimensional reconstructions from underwater robotic surveys," *Journal of Field Robotics*, vol. 27, no. 1, pp. 21–51, 2010.

[111] P. Jasiobedzki, C. Dimas, and D. Lim, "Underwater 3d modelling and photosynthetic life detection," in *Oceans, 2012*, Oct 2012, pp. 1–9.

[112] J. Aulinas, M. Carreras, X. Llado, J. Salvi, R. Garcia, R. Prados, and Y. Petillot, "Feature extraction for underwater visual slam," in *OCEANS, 2011 IEEE - Spain*, June 2011, pp. 1–7.

[113] J. Aulinas, Y. R. Petillot, X. Lladó, J. Salvi, and R. Garcia, "Vision-based underwater slam for the sparus auv," in *International Conference on Computer and IT Applications in the Maritime Industries (COMPIT)*, 2011, pp. 171–181.

[114] J.-Y. Bouguet. (2013, December) Camera calibration toolbox for matlab. [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc/

[115] B. Horn, "Closed-form solution of absolute orientation using unit quaternions," *JOSA A*, vol. 4, no. 4, pp. 629–642, 1987.

[116] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.

[117] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "g^2o: A general framework for graph optimization," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 3607–3613.

[118] OpenCV. (2013, December) OpenCV 2.4.8. [Online]. Available: http://www.opencv.org/

[119] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[120] J. Mallon and P. F. Whelan, "Calibration and removal of lateral chromatic aberration in images," *Pattern Recognition Letters*, vol. 28, no. 1, pp. 125 – 135, 2007.

[121] S. B. Kang, "Automatic removal of chromatic aberration from a single image," in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, June 2007, pp. 1–8.

[122] V. Kaufmann and R. Ladstädter, "Elimination of color fringes in digital photographs caused by lateral chromatic aberration," in *Proceedings of the XX International Symposium CIPA*, vol. 26, 2005, pp. 403–408.

[123] T. Boult and G. Wolberg, "Correcting chromatic aberrations using image warping," in *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR '92., 1992 IEEE Computer Society Conference on*, Jun 1992, pp. 684–687.

[124] J. Chang, H. Kang, and M. G. Kang, "Correction of axial and lateral chromatic aberration with false color filtering," *IEEE Transactions on Image Processing*, vol. 22, no. 3, pp. 1186–1198, March 2013.

[125] J. Góźdź. (2009) Dcb demosaicing. [Online]. Available: http://www.linuxphoto.org/html/dcb.html

[126] G. Horvath. (2010) Rawtherapee - demosaicing algorithms. [Online]. Available: http://rawtherapee.com

[127] Eigen. (2013, July) Eigen 3.2. [Online]. Available: http://eigen.tuxfamily.org/

[128] P. Jasiobedzki, S. Se, M. Bondy, and R. Jakola, "Underwater 3d mapping and pose estimation for rov operations," in *OCEANS 2008*, Sept 2008, pp. 1–6.

[129] C. Beall, B. Lawrence, V. Ila, and F. Dellaert, "3d reconstruction of underwater structures," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, Oct 2010, pp. 4418–4423.

[130] Y. Kwon, "Object plane deformation due to refraction in two-dimensional underwater motion analysis," *Journal of applied biomechanics*, vol. 15, no. 4, pp. 396–403, 1999.

[131] P. Ozog and R. M. Eustice, "On the importance of modeling camera calibration uncertainty in visual slam," in *Proceedings of the IEEE International Conference on Robotics and Automation*, Karlsruhe, Germany, May 2013, pp. 3762–3769.

[132] R. Eustice, H. Singh, J. Leonard, M. Walter, and R. Ballard, "Visually navigating the rms titanic with slam information filters," in *Proc. Robotics: Science and Systems*. Cambridge, MA, 2005.

[133] H. Bulow, I. Sokolovski, M. Pfingsthorn, and A. Birk, "Underwater stereo data acquisition and 3d registration with a spectral method," in *OCEANS - Bergen, 2013 MTS/IEEE*, June 2013, pp. 1–7.

[134] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[135] R. Ferreira, J. a. P. Costeira, and J. a. A. Santos, "Stereo reconstruction of a submerged scene," in *Proceedings of the Second Iberian conference on Pattern Recognition and Image Analysis - Volume Part I*, ser. IbPRIA'05. Berlin, Heidelberg: Springer-Verlag, 2005, pp. 102–109.

[136] R. Schattschneider, G. Maurino, and W. Wang, "Towards stereo vision slam based pose estimation for ship hull inspection," in *OCEANS 2011*, sept. 2011, pp. 1 –8.

[137] J. Shi and C. Tomasi, "Good features to track," in *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*, Jun 1994, pp. 593–600.

[138] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[139] R. Schattschneider and R. Green, "Enhanced ransac sampling based on non-repeating combinations," in *Proceedings of the 27th Conference on Image and Vision Computing New Zealand*, ser. IVCNZ '12. New York, NY, USA: ACM, 2012, pp. 198–203.

[140] D. Lowe, "Object recognition from local scale-invariant features," in *International Conference on Computer Vision (ICCV)*. Published by the IEEE Computer Society, 1999, p. 1150.

[141] M. Kaess, A. Ranganathan, and F. Dellaert, "isam: Incremental smoothing and mapping," *Robotics, IEEE Transactions on*, vol. 24, no. 6, pp. 1365 –1378, dec. 2008.

[142] G. Welch and G. Bishop, "An introduction to the kalman filter," University of North Carolina, Chapel Hill, NC, USA, Tech. Rep., 1995.

[143] A. Davison, I. Reid, N. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, p. 1052, 2007.

[144] T. Bailey, "Mobile robot localisation and mapping in extensive outdoor environments," Ph.D. dissertation, Australian Centre for Field Robotics, Department of Aerospace, Mechanical and Mechatronic Engineering, The University of Sydney, 2002.

[145] J.-L. Blanco, "Derivation and implementation of a full 6D EKF-based solution to bearing-range SLAM," University of Málaga, Tech. Rep., 2008.

[146] Lyttelton Port Company. (2010, March) Lyttelton port of christchurch. [Online]. Available: http://www.lpc.co.nz/

[147] MRPT. (2014, Feburary) The Mobile Robot Programming Toolkit 1.1.0. [Online]. Available: http://www.mrpt.org/

[148] J.-L. Blanco, "A tutorial on se(3)transformation parameterizations and on-manifold optimization," University of Málaga, Tech. Rep., 2012.

[149] R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," *Autonomous robot vehicles*, vol. 1, pp. 167–193, 1990.

Publications

1. R. Schattschneider, G. Maurino, and W. Wang, "Towards stereo vision slam based pose estimation for ship hull inspection," in *OCEANS 2011*, sept. 2011, pp. 1-8.
2. R. Schattschneider and R. Green, "Enhanced RANSAC sampling based on non-repeating combinations," in *Proceedings of the 27th Conference on Image and Vision Computing New Zealand*, 2012, pp. 198-203.